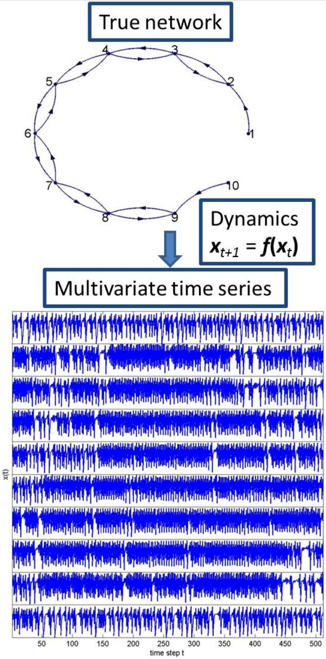
Time Series and Complex Networks

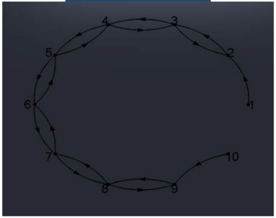
Kugiumtzis Dimitris

Department of Electrical and Computer Engineering,
Faculty of Engineering, Aristotle University of Thessaloniki, Greece
e-mail: dkugiu@auth.gr http:\\users.auth.gr\dkugiu

25th Summer School and Conference on Dynamical Systems and Complexity
Demokritos, 13 July 2017

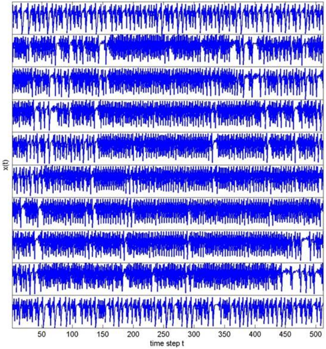
Complex networks from multivariate time series



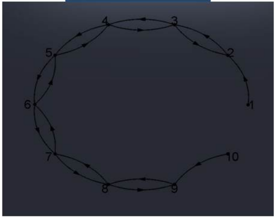


Complex networks from multivariate time series

Multivariate time series

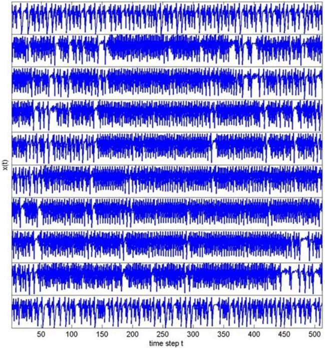


[redacted]



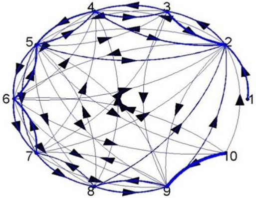
Complex networks from multivariate time series

Multivariate time series

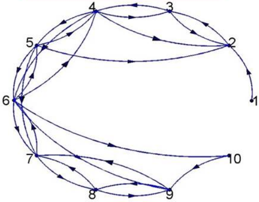


Estimated network

weighted connections



binary connections



Outline

Outline

1. Dependence measures in univariate time series

Outline

1. Dependence measures in univariate time series
2. Interdependence in multivariate time series

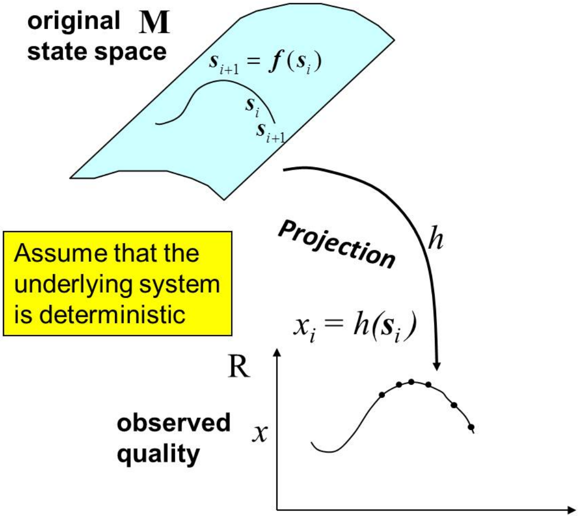
Outline

1. Dependence measures in univariate time series
2. Interdependence in multivariate time series
3. Complex networks from multivariate time series

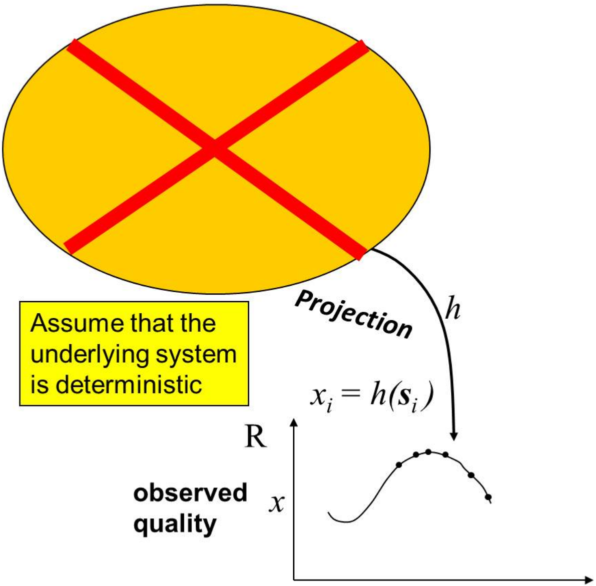
Outline

1. Dependence measures in univariate time series
2. Interdependence in multivariate time series
3. Complex networks from multivariate time series
4. High-dimensional time series: Implications and solutions

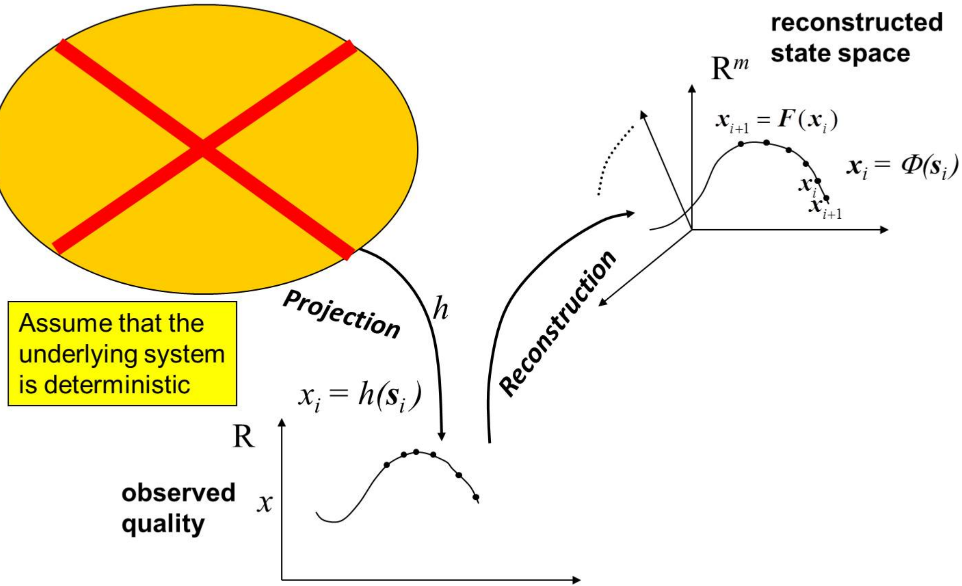
State space reconstruction (embedding)



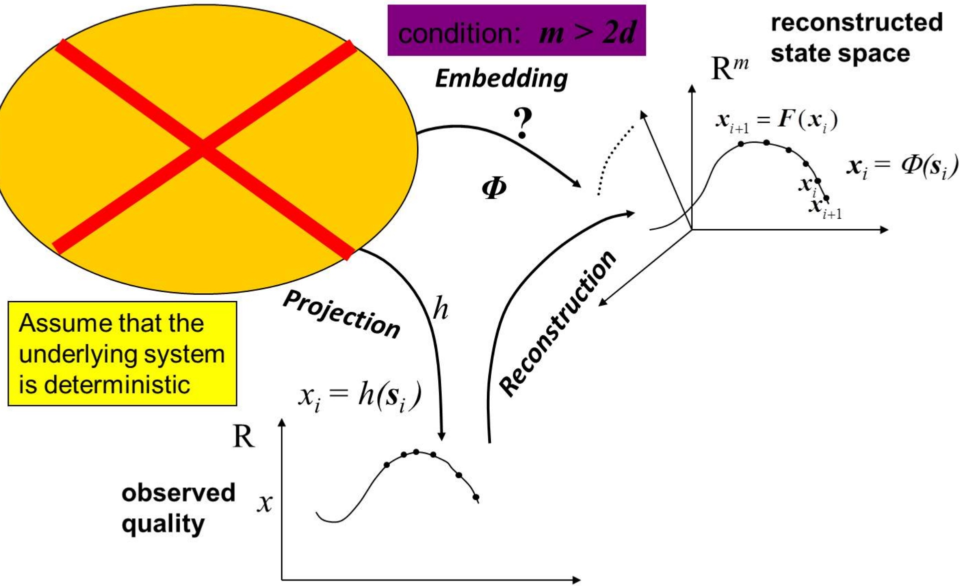
State space reconstruction (embedding)



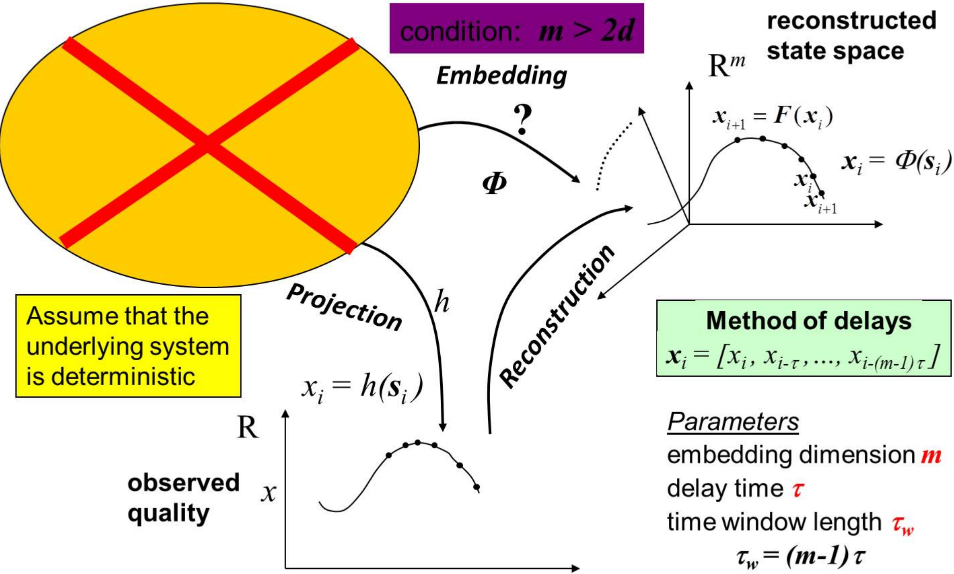
State space reconstruction (embedding)



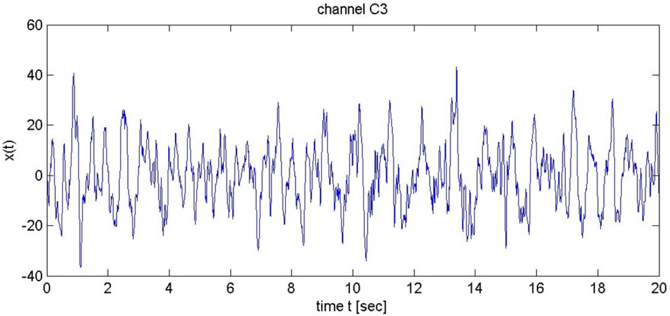
State space reconstruction (embedding)



State space reconstruction (embedding)



... Y_{t-5} Y_{t-4} Y_{t-3} Y_{t-2} Y_{t-1} Y_t Y_{t+1}



... $t-5$ $t-4$ $t-3$ $t-2$ $t-1$ t $t+1$

A single observable

Observation of an underlying system variable, $y_t = h(\boldsymbol{s}_t)$

… Y_{t-5} Y_{t-4} Y_{t-3} Y_{t-2} Y_{t-1} Y_t Y_{t+1}

Underlying dynamical system

$\boldsymbol{s}_{t+1} = \boldsymbol{f}(\boldsymbol{s}_t)$ or $\dot{\boldsymbol{s}} = \boldsymbol{f}(\boldsymbol{s})$, $\boldsymbol{s}_t \in \mathrm{R}^d$

or stochastic process

$\boldsymbol{s}_{t+1} = \boldsymbol{f}(\boldsymbol{s}_t, \boldsymbol{\varepsilon}_t)$ or $\dot{\boldsymbol{s}} = \boldsymbol{f}(\boldsymbol{s}, \boldsymbol{\varepsilon})$, $\boldsymbol{s}_t \in \mathrm{R}^d$

Reconstruction of the dynamics

$y_{t+1} = F(\boldsymbol{y}_t) = \mathrm{F}(y_t, y_{t-1}, \ldots, y_{t-m+1})$ $\quad y_t \in \mathrm{R}$

or

$y_{t+1} = F(\boldsymbol{y}_t) = \mathrm{F}(y_t, y_{t-\tau}, \ldots, y_{t-(m-1)\tau})$

delay embedding (uniform)

m : embedding dimension

τ : delay time

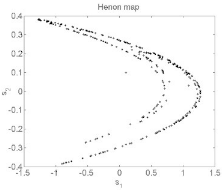
$\boldsymbol{s}_t$: state variable, e.g. representing the brain activity

y_t : observed dynamics, e.g. EEG

… $t-5$ $t-4$ $t-3$ $t-2$ $t-1$ t $t+1$

$y_t = s_t$

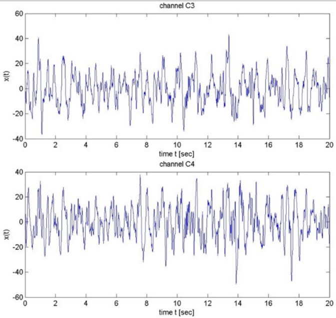
$s_t = 1 - 1.4 s_{t-1}^{\,2} + 0.3 s_{t-2}$

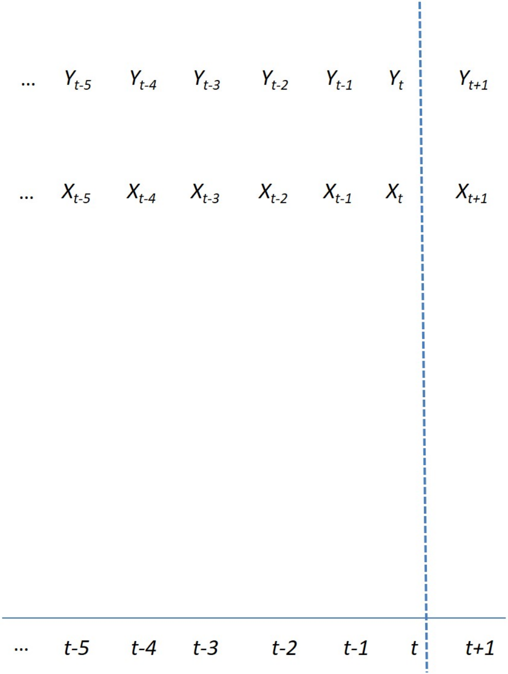


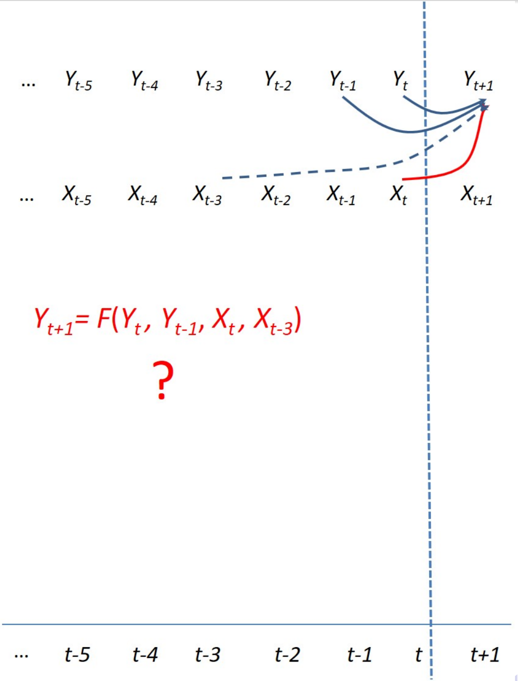
$\ldots \quad Y_{t-5} \quad Y_{t-4} \quad Y_{t-3} \quad Y_{t-2} \quad Y_{t-1} \quad Y_t \quad Y_{t+1}$

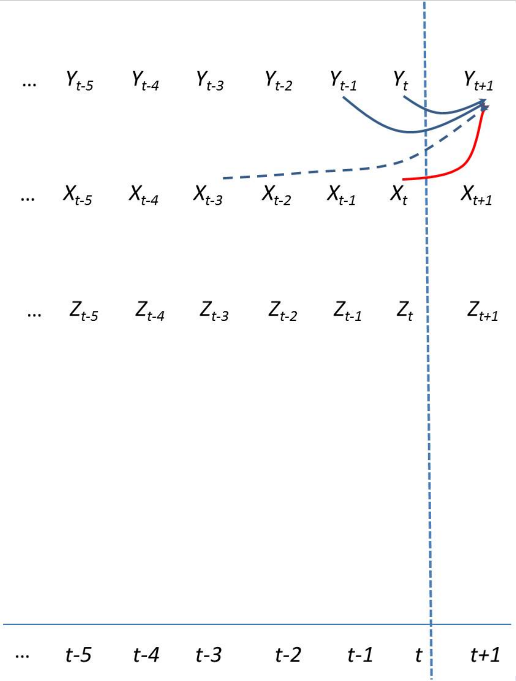
$Y_{t+1} = F(Y_t, Y_{t-1})$

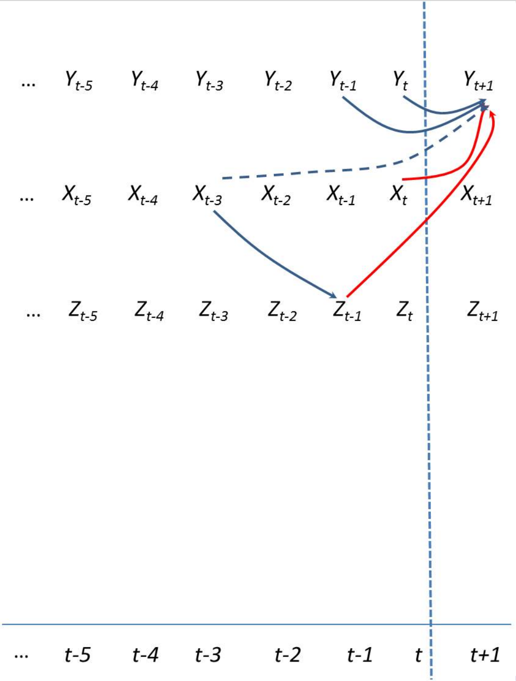
$\ldots \quad t-5 \quad t-4 \quad t-3 \quad t-2 \quad t-1 \quad t \quad t+1$

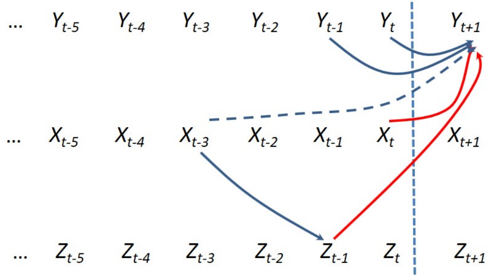

channel C3
x(t)
time t [sec]
channel C4
x(t)
time t [sec]


... Y_{t-5} Y_{t-4} Y_{t-3} Y_{t-2} Y_{t-1} Y_t Y_{t+1}
... X_{t-5} X_{t-4} X_{t-3} X_{t-2} X_{t-1} X_t X_{t+1}
... t-5 t-4 t-3 t-2 t-1 t t+1

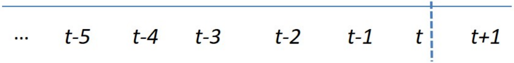

… Y_{t-5} Y_{t-4} Y_{t-3} Y_{t-2} Y_{t-1} Y_t Y_{t+1}
… X_{t-5} X_{t-4} X_{t-3} X_{t-2} X_{t-1} X_t X_{t+1}
$Y_{t+1} = F(Y_t, Y_{t-1}, X_t, X_{t-3})$
?
… t-5 t-4 t-3 t-2 t-1 t t+1

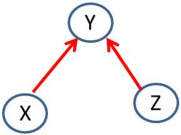

... Y_{t-5} Y_{t-4} Y_{t-3} Y_{t-2} Y_{t-1} Y_t Y_{t+1}
... X_{t-5} X_{t-4} X_{t-3} X_{t-2} X_{t-1} X_t X_{t+1}
... Z_{t-5} Z_{t-4} Z_{t-3} Z_{t-2} Z_{t-1} Z_t Z_{t+1}
... t-5 t-4 t-3 t-2 t-1 t t+1


… Y_{t-5} Y_{t-4} Y_{t-3} Y_{t-2} Y_{t-1} Y_t Y_{t+1}
… X_{t-5} X_{t-4} X_{t-3} X_{t-2} X_{t-1} X_t X_{t+1}
… Z_{t-5} Z_{t-4} Z_{t-3} Z_{t-2} Z_{t-1} Z_t Z_{t+1}
… t-5 t-4 t-3 t-2 t-1 t t+1

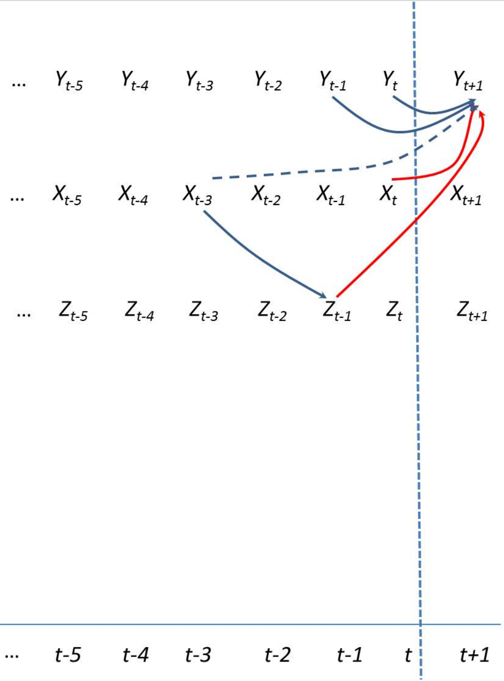


$$Y_{t+1}= F(Y_t, Y_{t-1}, X_t, Z_{t-1})$$

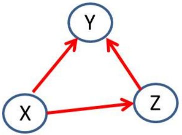


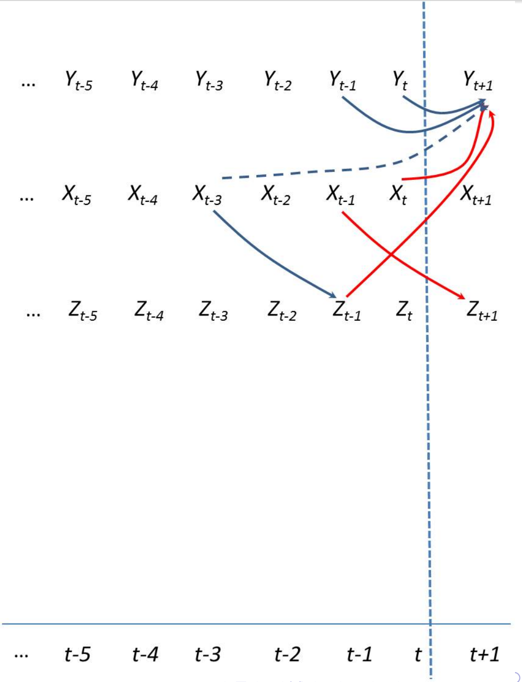


Y
X
Z



... Y_{t-5} Y_{t-4} Y_{t-3} Y_{t-2} Y_{t-1} Y_t Y_{t+1}
... X_{t-5} X_{t-4} X_{t-3} X_{t-2} X_{t-1} X_t X_{t+1}
... Z_{t-5} Z_{t-4} Z_{t-3} Z_{t-2} Z_{t-1} Z_t Z_{t+1}
... t-5 t-4 t-3 t-2 t-1 t t+1

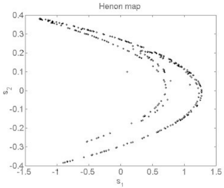

Y
X
Z

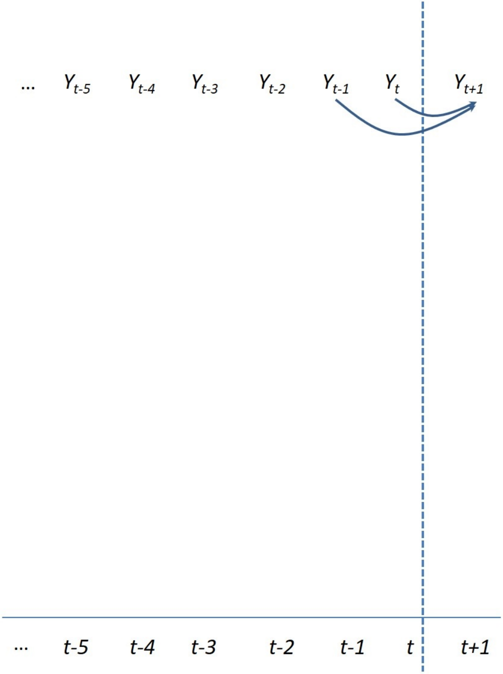

... Y_{t-5} Y_{t-4} Y_{t-3} Y_{t-2} Y_{t-1} Y_t Y_{t+1}
... X_{t-5} X_{t-4} X_{t-3} X_{t-2} X_{t-1} X_t X_{t+1}
... Z_{t-5} Z_{t-4} Z_{t-3} Z_{t-2} Z_{t-1} Z_t Z_{t+1}
... t-5 t-4 t-3 t-2 t-1 t t+1

1. Dependence measures in univariate time series
2. Interdependence in multivariate time series
3. Complex networks from multivariate time series
4. High-dimensional time series: Implications and solutions

$y_t = s_t$

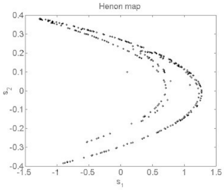
$s_t = 1 - 1.4 s_{t-1}^{\,2} + 0.3 s_{t-2}$





$y_t = s_t$

$s_t = 1 - 1.4 s_{t-1}^{\,2} + 0.3 s_{t-2}$



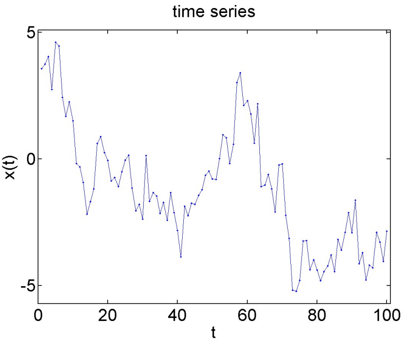
… Y_{t-5} Y_{t-4} Y_{t-3} Y_{t-2} Y_{t-1} Y_t Y_{t+1}

Are Y_t and Y_{t-1} correlated?
Are Y_t and Y_{t-2} correlated?

… $t-5$ $t-4$ $t-3$ $t-2$ $t-1$ t $t+1$

Are X_t and X_{t-1} linearly correlated?

Are X_t and X_{t-2} linearly correlated?

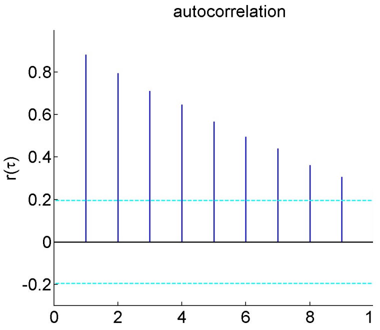


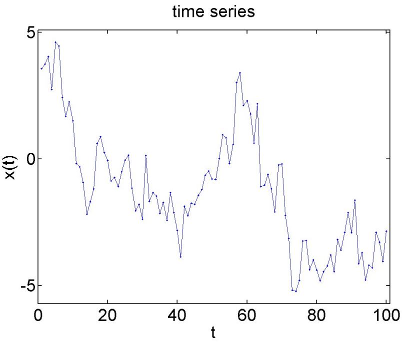
Autocorrelation $r(\tau) = r(X_t; X_{t-\tau})$

Are X_t and X_{t-1} linearly correlated? $r(1) = r(X_t; X_{t-1}) \neq 0$? Yes

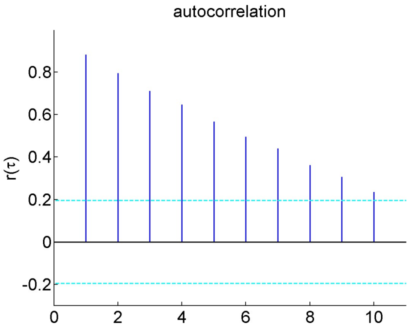
Are X_t and X_{t-2} linearly correlated? $r(2) = r(X_t; X_{t-2}) \neq 0$? Yes

$$r(\tau) = r(X_t; X_{t-\tau}) = \frac{1}{n-\tau} \sum_{t=\tau+1}^{n} (x_t - \bar{x})(x_{t-\tau} - \bar{x}) / s_X^2$$





Are X_t and X_{t-2} directly linearly correlated?

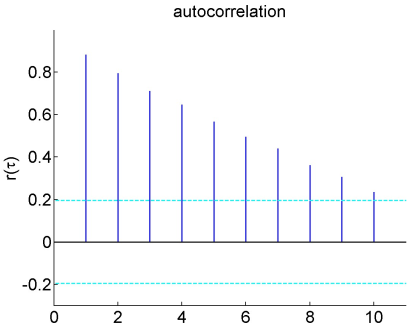


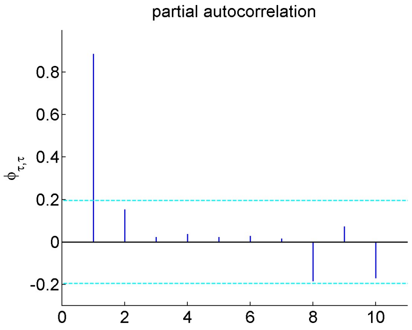
Partial autocorrelation $\phi_{\tau,\tau} = r(X_t; X_{t-\tau} | X_{t-1}, \ldots, X_{t-\tau+1})$

Are X_t and X_{t-2} directly linearly correlated?

Are X_t and X_{t-2} linearly correlated given X_{t-1}?

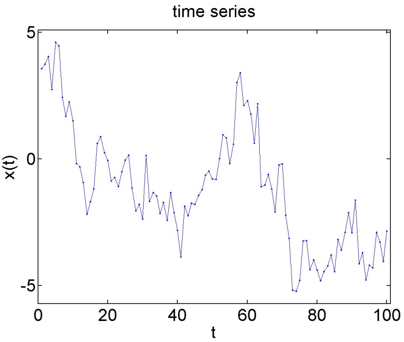
$$r(X_t; X_{t-2} | X_{t-1}) \neq 0? \qquad \text{No}$$





Are X_t and X_{t-1} linearly and nonlinearly correlated?

Are X_t and X_{t-2} linearly and nonlinearly correlated?



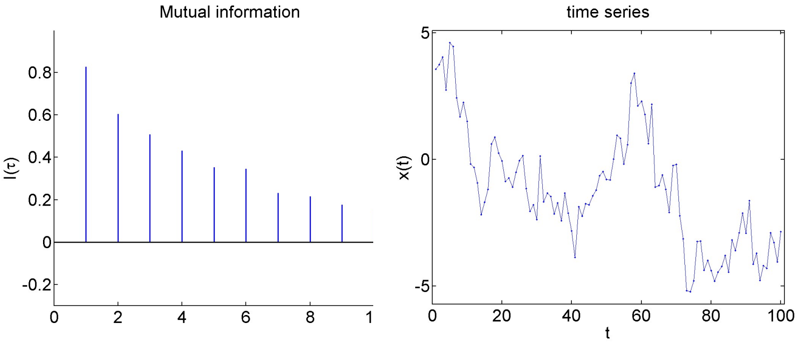
Mutual information $I(\tau) = I(X_t; X_{t-\tau})$

Are X_t and X_{t-1} linearly and nonlinearly correlated? $I(X_t; X_{t-1}) \neq 0$? Yes

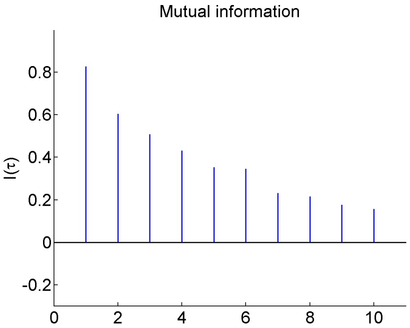
Are X_t and X_{t-2} linearly and nonlinearly correlated? $I(X_t; X_{t-2}) \neq 0$? Yes

$$I(\tau) = I(X_t, X_{t-\tau}) = H(X_t) + H(X_{t-\tau}) - H(X_t, X_{t-\tau})$$
$$= \sum_{x_t, x_{t-\tau}} p_{X_t X_{t-\tau}}(x_t, x_{t-\tau}) \log \frac{p_{X_t X_{t-\tau}}(x_t, x_{t-\tau})}{p_{X_t}(x_t) p_{X_{t-\tau}}(x_{t-\tau})}$$

$$H(X) = -\sum_x p(x) \log p(x)$$



Are X_t and X_{t-2} directly linearly and nonlinearly correlated?

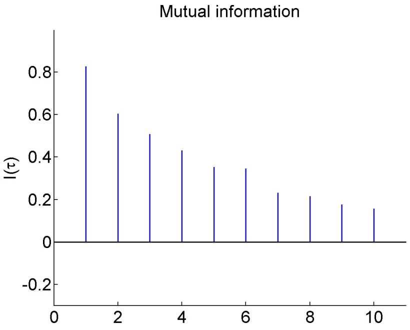


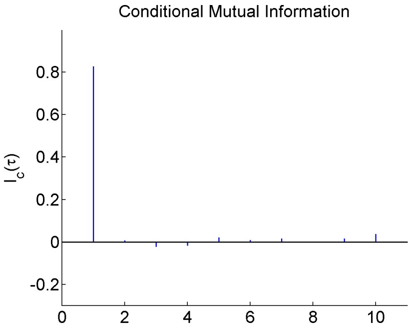
Conditional mutual information $I_c(\tau) = I(X_t; X_{t-\tau} | X_{t-1}, \ldots, X_{t-\tau+1})$

Are X_t and X_{t-2} directly linearly and nonlinearly correlated?

Are X_t and X_{t-2} linearly and nonlinearly correlated given X_{t-1}?

$$I(X_t; X_{t-2} | X_{t-1}) \neq 0? \qquad \text{No}$$



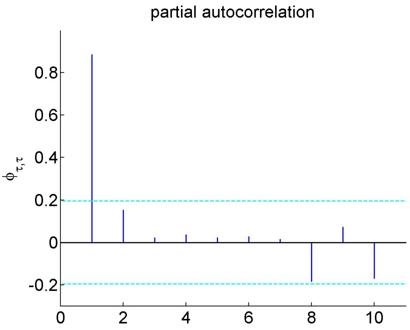


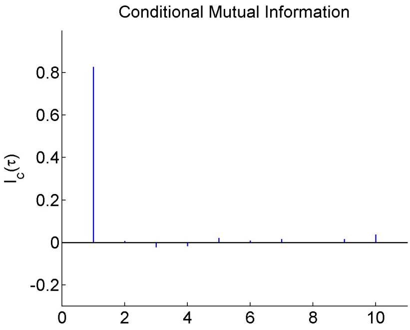
Are X_t and X_{t-2} directly linearly correlated?

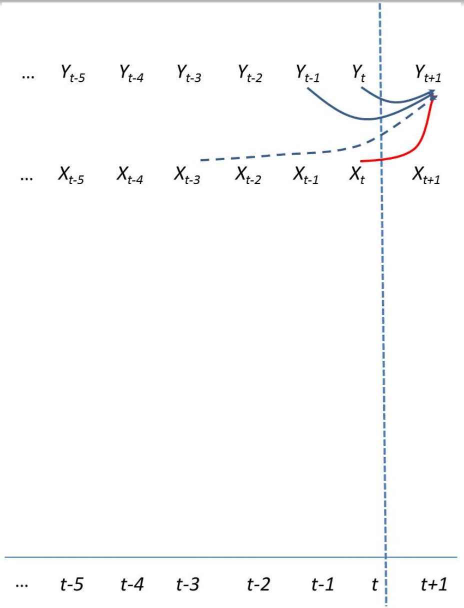
$$r(X_t; X_{t-2}|X_{t-1}) \neq 0? \qquad \text{No}$$

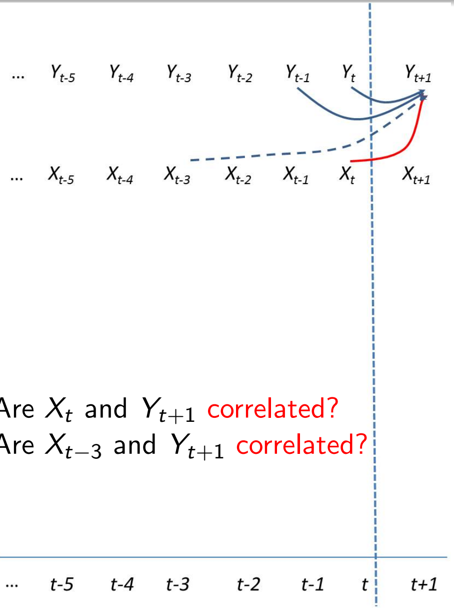
Are X_t and X_{t-2} directly linearly or/and nonlinearly correlated?

$$I(X_t; X_{t-2}|X_{t-1}) \neq 0? \qquad \text{No}$$

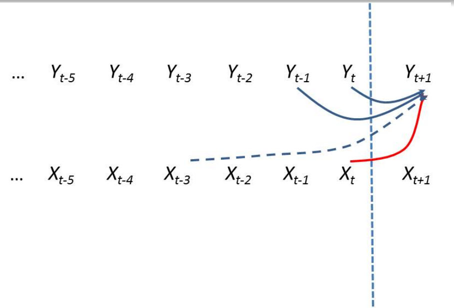




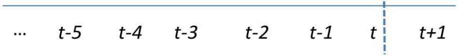

... Y_{t-5} Y_{t-4} Y_{t-3} Y_{t-2} Y_{t-1} Y_t Y_{t+1}
... X_{t-5} X_{t-4} X_{t-3} X_{t-2} X_{t-1} X_t X_{t+1}
... t-5 t-4 t-3 t-2 t-1 t t+1

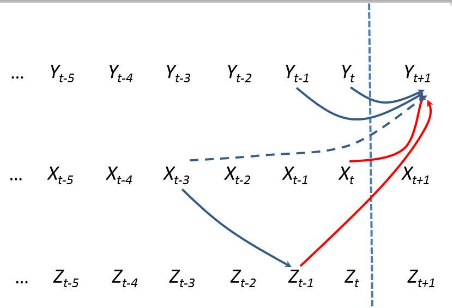


Are X_t and Y_{t+1} correlated?
Are X_{t-3} and Y_{t+1} correlated?

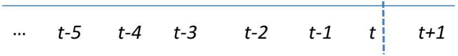


Are X_t and Y_{t+1} correlated?
Are X_{t-3} and Y_{t+1} correlated?



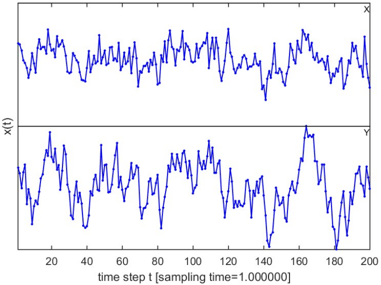


Are X_t and Y_{t+1} correlated?
Are X_{t-3} and Y_{t+1} correlated?



Are X_t and Y_{t+1} linearly correlated?

Are X_{t-3} and Y_{t+1} linearly correlated?

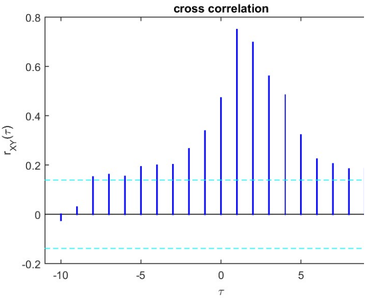


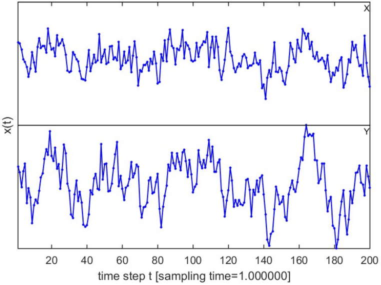
Cross-correlation $r_{XY}(\tau) = r(X_t; Y_{t+\tau})$

Are X_t and Y_{t+1} linearly correlated? $r(1) = r(X_t; Y_{t+1}) \neq 0$? Yes

Are X_{t-3} and Y_{t+1} linearly correlated? $r(4) = r(X_t; Y_{t+4}) \neq 0$? Yes

$$r_{XY}(\tau) = r(X_t; Y_{t+\tau}) = \frac{1}{n-\tau} \sum_{t=1}^{n-\tau} (x_t - \bar{x})(y_{t+\tau} - \bar{y})/(s_X s_Y)$$



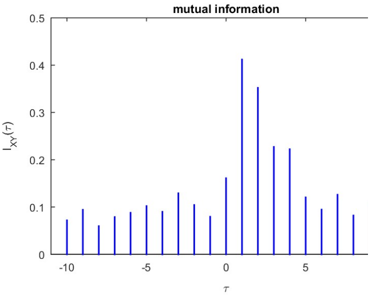


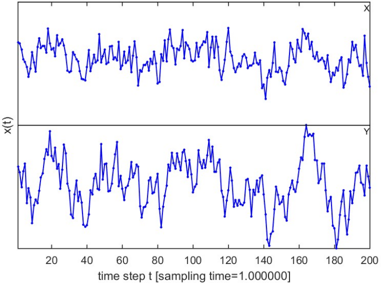
Cross-mutual information $I_{XY}(\tau) = I(X_t; Y_{t+\tau})$

Are X_t and Y_{t+1} linearly and nonlinearly correlated? Yes

Are X_{t-3} and Y_{t+1} linearly and nonlinearly correlated? Yes

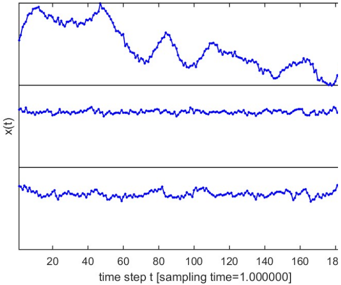
$$I_{XY}(\tau) = I(X_t, Y_{t+\tau}) = H(X_t) + H(Y_{t+\tau}) - H(X_t, Y_{t+\tau})$$
$$= \sum_{x_t, y_{t+\tau}} p_{X_t Y_{t+\tau}}(x_t, y_{t+\tau}) \log \frac{p_{X_t Y_{t+\tau}}(x_t, y_{t+\tau})}{p_{X_t}(x_t) p_{Y_{t+\tau}}(y_{t+\tau})}$$

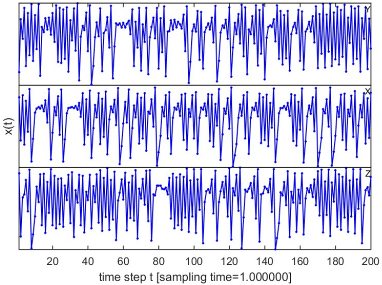




Are X_t and Y_{t+1} directly correlated?

Are X_{t-3} and Y_{t+1} directly correlated?

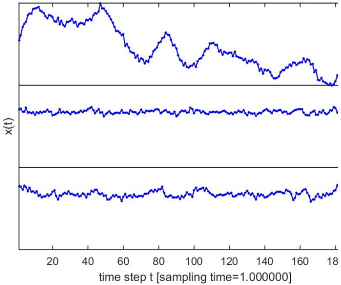


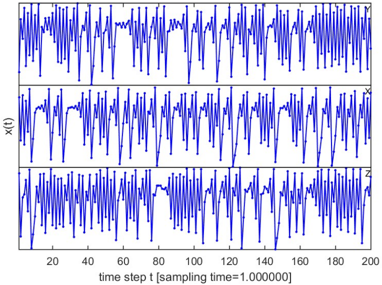


Are X_t and Y_{t+1} directly correlated?

Are X_{t-3} and Y_{t+1} directly correlated?

Partial Cross-correlation $r_{XY|Z}(0) = r(X_t; Y_t|Z_t)$

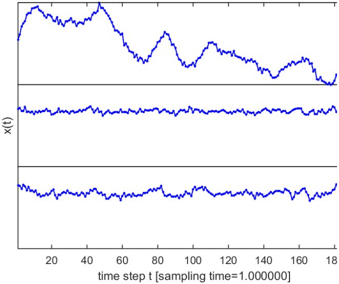


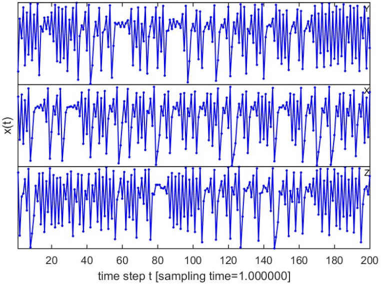


Are X_t and Y_{t+1} directly correlated?

Are X_{t-3} and Y_{t+1} directly correlated?

Partial Cross-correlation $r_{XY|Z}(0) = r(X_t; Y_t|Z_t)$ **lags?**



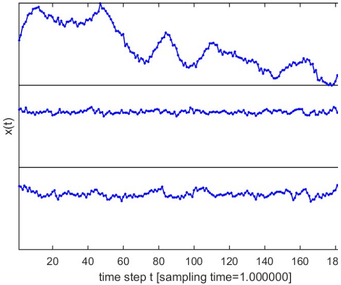


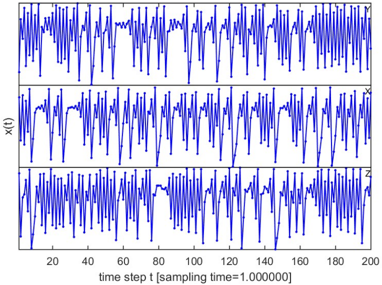
Are X_t and Y_{t+1} directly correlated?

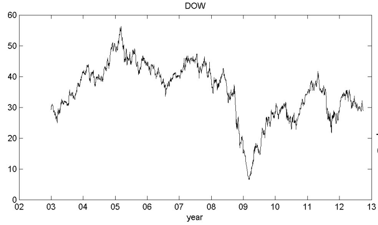
Are X_{t-3} and Y_{t+1} directly correlated?

Partial Cross-correlation $r_{XY|Z}(0) = r(X_t; Y_t|Z_t)$ **lags?**

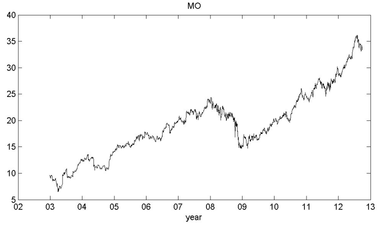
Partial Cross-mutual information $I_{XY|Z}(0) = I(X_t; Y_t|Z_t)$ **lags?**

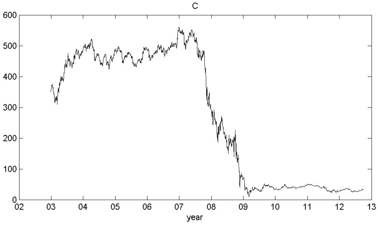


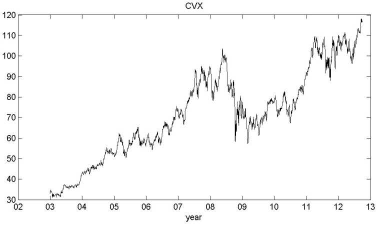




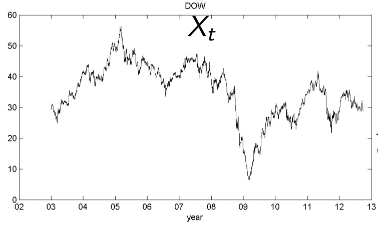
correlation
?



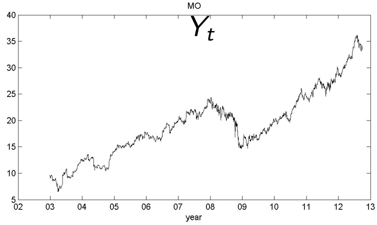


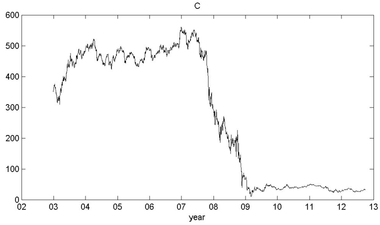


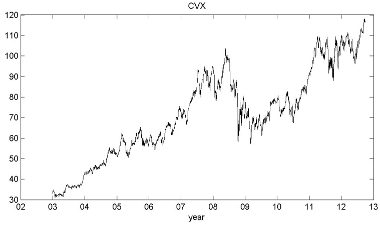
$$r(X_t; Y_t)$$



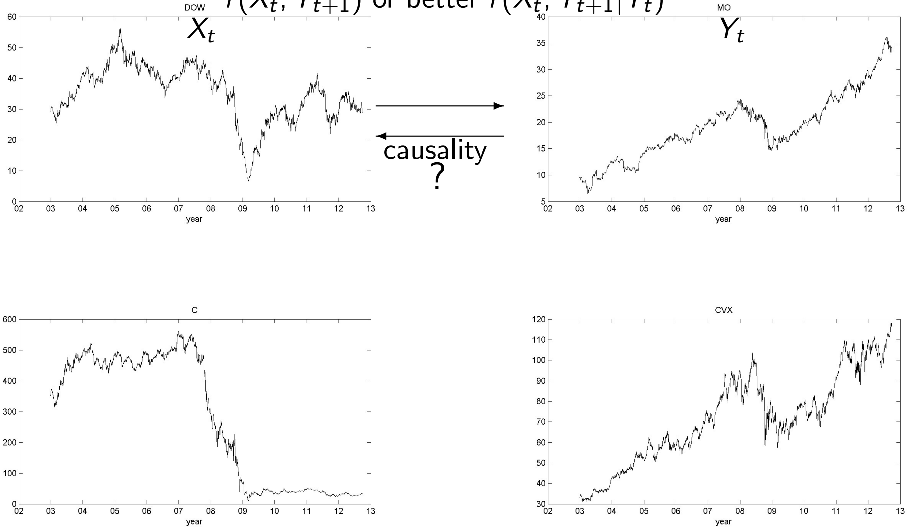
correlation
?



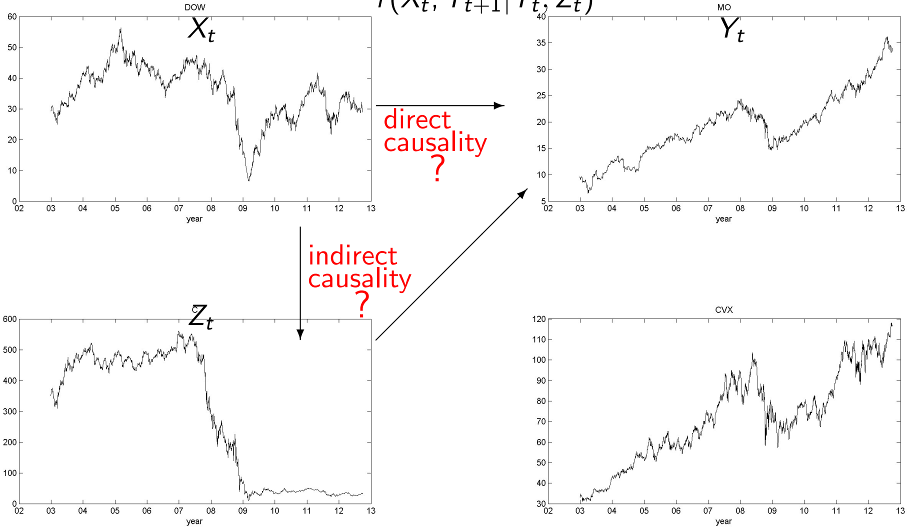




$r(X_t; Y_{t+1})$ or better $r(X_t; Y_{t+1} | Y_t)$



$$r(X_t; Y_{t+1} | Y_t, Z_t)$$


DOW
X_t
direct
causality
?
MO
Y_t
indirect
causality
?
C
Z_t
CVX
year

$r_{XY}(0) \neq 0$ or $I_{XY}(0) \neq 0$: $\Longrightarrow$ correlation of x_t and y_t

$\Longrightarrow$ systems X and Y are correlated, $X \sim Y$

$r_{XY}(0) \neq 0$ or $I_{XY}(0) \neq 0$: $\quad \Longrightarrow$ correlation of x_t and y_t

$\Longrightarrow$ systems X and Y are correlated, $\quad X \sim Y$

$r_{XY}(\tau) \neq 0$ or $I_{XY}(\tau) \neq 0$: $\quad \Longrightarrow$ correlation of x_t and $y_{t+\tau}$

$\Longrightarrow$ X effects the future of Y, $\quad X \rightarrow Y$

$r_{XY}(0) \neq 0$ or $I_{XY}(0) \neq 0$: $\quad \Longrightarrow$ correlation of x_t and y_t

$\Longrightarrow$ systems X and Y are correlated, $\quad X \sim Y$

$r_{XY}(\tau) \neq 0$ or $I_{XY}(\tau) \neq 0$: $\quad \Longrightarrow$ correlation of x_t and $y_{t+\tau}$

$\Longrightarrow$ X effects the future of Y, $\quad X \rightarrow Y$

$r_{XY}(-\tau) \neq 0$ or $I_{XY}(-\tau) \neq 0$: $\quad \Longrightarrow \quad Y \rightarrow X$

$r_{XY}(0) \neq 0$ or $I_{XY}(0) \neq 0$: $\quad \Longrightarrow$ correlation of x_t and y_t

$\Longrightarrow$ systems X and Y are correlated, $\quad X \sim Y$

$r_{XY}(\tau) \neq 0$ or $I_{XY}(\tau) \neq 0$: $\quad \Longrightarrow$ correlation of x_t and $y_{t+\tau}$

$\Longrightarrow$ X effects the future of Y, $\quad X \rightarrow Y$

$r_{XY}(-\tau) \neq 0$ or $I_{XY}(-\tau) \neq 0$: $\quad \Longrightarrow \quad Y \rightarrow X$

Thus $r_{XY}(\tau)$ and $I_{XY}(\tau)$ indicate the direction of interaction.

$r_{XY}(0) \neq 0$ or $I_{XY}(0) \neq 0$: $\Longrightarrow$ correlation of x_t and y_t

$\Longrightarrow$ systems X and Y are correlated, $X \sim Y$

$r_{XY}(\tau) \neq 0$ or $I_{XY}(\tau) \neq 0$: $\Longrightarrow$ correlation of x_t and $y_{t+\tau}$

$\Longrightarrow$ X effects the future of Y, $X \to Y$

$r_{XY}(-\tau) \neq 0$ or $I_{XY}(-\tau) \neq 0$: $\Longrightarrow$ $Y \to X$

Thus $r_{XY}(\tau)$ and $I_{XY}(\tau)$ indicate the direction of interaction.

Can they also be used as causality measures?

$r_{XY}(0) \neq 0$ or $I_{XY}(0) \neq 0$: $\quad \Longrightarrow$ correlation of x_t and y_t

$\Longrightarrow$ systems X and Y are correlated, $\quad X \sim Y$

$r_{XY}(\tau) \neq 0$ or $I_{XY}(\tau) \neq 0$: $\quad \Longrightarrow$ correlation of x_t and $y_{t+\tau}$

$\Longrightarrow$ X effects the future of Y, $\quad X \rightarrow Y$

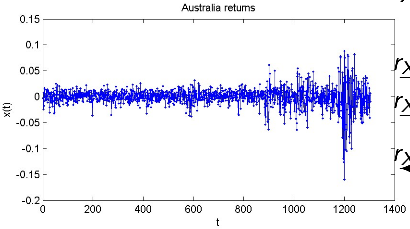
$r_{XY}(-\tau) \neq 0$ or $I_{XY}(-\tau) \neq 0$: $\quad \Longrightarrow \quad Y \rightarrow X$

Thus $r_{XY}(\tau)$ and $I_{XY}(\tau)$ indicate the direction of interaction.

Can they also be used as causality measures?
Not the most appropriate, but they have been used in many studies

Example: Returns for USA, UnitedKingdom, Greece and Australia.

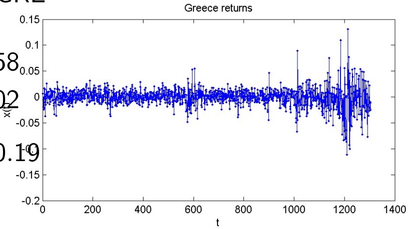
X:AUS, Y:GRE



$r_{XY}(0) = 0.58$

$r_{XY}(1) = 0.02$

$r_{XY}(-1) = 0.19$

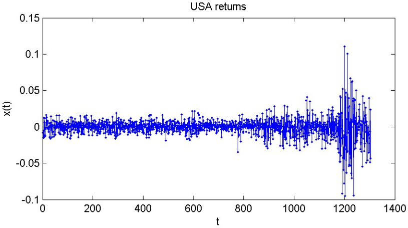


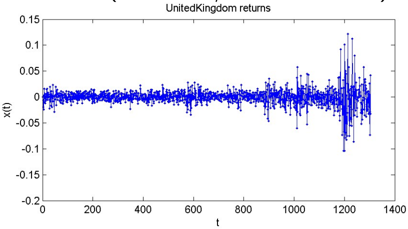
returns:

$x_t = \log(y_t) - \log(y_{t-1})$

Is the measure significant?

Can I draw a link? (directed/non-directed)

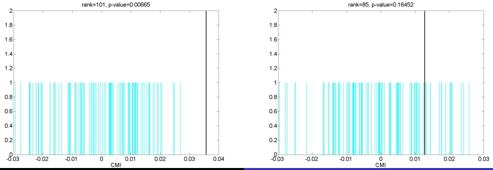




Significance randomization test for a correlation / causality measure q, $\mathrm{H}_0: q=0 \qquad \mathrm{H}_1: q \neq 0$

1. Generate M resampled (surrogate) time series, each by shifting the original observations with a random time step w:
   original time series: $\{x_t\} = \{x_1, x_2, \ldots, x_n\}$
   i-th surrogate time series:
   $\{x_t^{*i}\} = \{x_{w+1}, x_{w+2}, \ldots, x_n, x_1, \ldots, x_{w-1}, x_w\}$
2. Compute the statistic q on the original pair, q_0, and on the M surrogate pairs, $q_1, \ldots, q_M$,
   e.g. $q_0 \equiv r_{XY}(\tau) = \mathrm{Corr}(x_t, y_{t+\tau})$ and $q_i \equiv \mathrm{Corr}(x_t^{*i}, y_{t+\tau}^{*i})$
3. If q_0 is at the tails of the empirical null distribution formed by $q_1, \ldots, q_M$, reject H_0.
   Using rank ordering: for a two-sided test, the p-value of the test is

$$
\begin{array}{ll}
2\frac{r_{q_0}-0.326}{M+1+0.348} & \text{if} \quad r_{q_0} < \frac{M+1}{2} \\
2\left(1-\frac{r_{q_0}-0.326}{M+1+0.348}\right) & \text{if} \quad r_{q_0} \geq \frac{M+1}{2}
\end{array}
$$



Example: Returns for USA, UnitedKingdom, Greece and Australia.
Correlation matrix for delay 1, $r_{XY}(1)$

$$R(1) = \begin{bmatrix} & 0.382 & 0.333 & 0.596 \\ 0.049 & & 0.039 & 0.303 \\ 0.096 & 0.001 & & 0.190 \\ 0.031 & -0.001 & -0.021 & \end{bmatrix}$$

Randomization significance test for $r_{XY}(1)$ ($M = 1000$)

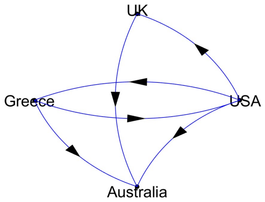
Matrix of p-values | Adjacency matrix

$$P(R(1)) = \begin{bmatrix} & 0.0013 & 0.0013 & 0.0033 \\ 0.0732 & & 0.1991 & 0.0013 \\ 0.0073 & 0.8901 & & 0.0033 \\ 0.2450 & 0.9760 & 0.4028 & \end{bmatrix} A = \begin{bmatrix} & 1 & 1 & 1 \\ 0 & & 0 & 1 \\ 1 & 0 & & 1 \\ 0 & 0 & 0 & \end{bmatrix}$$

For significance level, say $\alpha = 0.05$, there may be $p < \alpha$ more often than it should be due to multiple testing.
Correction with e.g. False Discovery Rate (FDR)

Example: Returns for USA, UnitedKingdom, Greece and Australia.
Correlation matrix for delay 1, $r_{XY}(1)$

$$R(1) = \begin{bmatrix} & 0.382 & 0.333 & 0.596 \\ 0.049 & & 0.039 & 0.303 \\ 0.096 & 0.001 & & 0.190 \\ 0.031 & -0.001 & -0.021 & \end{bmatrix}$$



Randomization significance test for $r_{XY}(1)$ ($M = 1000$)

Matrix of p-values

$$P(R(1)) = \begin{bmatrix} & 0.0013 & 0.0013 & 0.0033 \\ 0.0732 & & 0.1991 & 0.0013 \\ 0.0073 & 0.8901 & & 0.0033 \\ 0.2450 & 0.9760 & 0.4028 & \end{bmatrix}$$

Adjacency matrix

$$A = \begin{bmatrix} & 1 & 1 & 1 \\ 0 & & 0 & 1 \\ 1 & 0 & & 1 \\ 0 & 0 & 0 & \end{bmatrix}$$

For significance level, say $\alpha = 0.05$, there may be $p < \alpha$ more often than it should be due to multiple testing.
Correction with e.g. False Discovery Rate (FDR)

Linear causality measures (direct and indirect)

Idea of Granger causality $X \to Y$ [Granger 1969]:
predict Y better when including X in the regression model.

Granger Causality Index (GCI) [Brandt & Williams 2007]

Bivariate time series $\{x_t, y_t\}_{t=1}^{n}$
driving system: X, response system: Y

Model 1 (restricted, R, X absent in the model):

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + e_{R,t}$$

Model 2 (unrestricted, U, X present in the model):

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{p} b_i x_{t-i} + e_{U,t}$$

$$\mathrm{GCI}_{X\to Y} = \ln \frac{\mathrm{Var}(\hat{e}_{R,t})}{\mathrm{Var}(\hat{e}_{U,t})} \qquad \mathrm{GCI}_{X\to Y} > 0 \Rightarrow X \to Y \text{ holds}$$

Parametric significance test for GCI

$\text{GCI}_{X \to Y} > 0$? $\Rightarrow$ Significance test

If X does not Granger causes Y then the contribution of X-lags in the unrestricted model should be insignificant $\Rightarrow$
the terms of X should be insignificant

H_0: $b_i = 0$, for all $i = 1, \ldots, p$
H_1: $b_i \neq 0$, for any of $i = 1, \ldots, p$

Snedecor-Fisher test (F-test):

$$F = \frac{(\text{SSE}^R - \text{SSE}^U)/p}{\text{SSE}^U/\text{ndf}}$$

SSE: sum of squared errors
ndf: number of degrees of freedoms, $\text{ndf} = (n-p) - 2p$,
$n - p$: number of equations,
$2p$: number of coefficients in the U-model.

Conditional Granger Causality Index (CGCI)

K time series $\{x_t, y_t\}_{t=1}^n$ and $\{\mathbf{z}_t\}_{t=1}^n = \{z_{1,t}, z_{2,t}, \ldots, z_{K-2,t}\}_{t=1}^n$
driving system: X, response system: Y,
conditioning on system Z, $Z = \{Z_1, Z_2, \ldots, Z_{K-2}\}$

Model 1 (restricted, R, X absent in the model):

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{p} A_i \mathbf{z}_{t-i} + e_{R,t}$$

Model 2 (unrestricted, U, X present in the model):

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{p} b_i x_{t-i} + \sum_{i=1}^{p} A_i \mathbf{z}_{t-i} + e_{U,t}$$

$$\mathrm{CGCI}_{X \to Y|Z} = \ln \frac{\mathrm{Var}(\hat{e}_{R,t})}{\mathrm{Var}(\hat{e}_{U,t})}$$

Parametric significance test for CGCI

$\text{CGCI}_{X \to Y|Z} > 0$ **?** $\Rightarrow$ Significance test as for GCI

H_0: $b_i = 0$, for all $i = 1, \dots, p$
H_1: $b_i \neq 0$, for any of $i = 1, \dots, p$

$$F = \frac{(\text{SSE}^R - \text{SSE}^U)/p}{\text{SSE}^U/\text{ndf}}$$

$\text{ndf} = (n-p) - Kp$,
$n-p$: number of equations,
Kp: number of coefficients in the U-model.

Model order and embedding parameters

VAR model for Y

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{p} b_i x_{t-i} + e_{U,t}$$

Model order and embedding parameters

VAR model for Y

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{p} b_i x_{t-i} + e_{U,t}$$

$$y_{t+1} = \sum_{i=1}^{p} a_i y_{t-i+1} + \sum_{i=1}^{p} b_i x_{t-i+1} + e_{U,t+1}$$

y_{t+1} is given in terms of $\mathbf{y}_t = [y_t, y_{t-1}, \ldots, y_{t-p+1}]$ and $\mathbf{x}_t = [x_t, x_{t-1}, \ldots, x_{t-p+1}]$, $\quad y_{t+1} = \mathbf{F}(\mathbf{y}_t, \mathbf{x}_t)$

Model order and embedding parameters

VAR model for Y

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{p} b_i x_{t-i} + e_{U,t}$$

$$y_{t+1} = \sum_{i=1}^{p} a_i y_{t-i+1} + \sum_{i=1}^{p} b_i x_{t-i+1} + e_{U,t+1}$$

y_{t+1} is given in terms of $\mathbf{y}_t = [y_t, y_{t-1}, \ldots, y_{t-p+1}]$ and $\mathbf{x}_t = [x_t, x_{t-1}, \ldots, x_{t-p+1}]$, $\quad y_{t+1} = \mathbf{F}(\mathbf{y}_t, \mathbf{x}_t)$

Let the lag step be $\tau \geq 1 \quad \Rightarrow \quad \mathbf{y}_t = [y_t, y_{t-\tau}, \ldots, y_{t-(p-1)\tau}]$:

τ, p: embedding parameters (generally different for X and Y)

Model order and embedding parameters

VAR model for Y

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{p} b_i x_{t-i} + e_{U,t}$$

$$y_{t+1} = \sum_{i=1}^{p} a_i y_{t-i+1} + \sum_{i=1}^{p} b_i x_{t-i+1} + e_{U,t+1}$$

y_{t+1} is given in terms of $\mathbf{y}_t = [y_t, y_{t-1}, \ldots, y_{t-p+1}]$ and $\mathbf{x}_t = [x_t, x_{t-1}, \ldots, x_{t-p+1}]$, $\quad y_{t+1} = \mathbf{F}(\mathbf{y}_t, \mathbf{x}_t)$

Let the lag step be $\tau \geq 1 \quad \Rightarrow \quad \mathbf{y}_t = [y_t, y_{t-\tau}, \ldots, y_{t-(p-1)\tau}]$:

τ, p: embedding parameters (generally different for X and Y)

State space reconstruction:

$\mathbf{x}_t = [x_t, x_{t-\tau_x}, \ldots, x_{t-(m_x-1)\tau_x}]'$, embedding parameters: m_x, τ_x

$\mathbf{y}_t = [y_t, y_{t-\tau_y}, \ldots, y_{t-(m_y-1)\tau_y}]'$, embedding parameters: m_y, τ_y

y_{t+1}: future state of Y

Nonlinear causality measures (direct and indirect)

Transfer Entropy (TE) [Schreiber, 2000]

Measure the effect of X on Y at one time step ahead, accounting (conditioning) for the effect from its own current state

$$
\begin{aligned}
\mathrm{TE}_{X\to Y} &= I(y_{t+1};\mathbf{x}_t|\mathbf{y}_t) \\
&= H(\mathbf{x}_t,\mathbf{y}_t) - H(y_{t+1},\mathbf{x}_t,\mathbf{y}_t) + H(y_{t+1},\mathbf{y}_t) - H(\mathbf{y}_t) \\
&= \sum p(y_{t+1},\mathbf{x}_t,\mathbf{y}_t)\log\frac{p(y_{t+1}|\mathbf{x}_t,\mathbf{y}_t)}{p(y_{t+1}|\mathbf{y}_t)}
\end{aligned}
$$

Nonlinear causality measures (direct and indirect)

Transfer Entropy (TE) [Schreiber, 2000]

Measure the effect of X on Y at one time step ahead, accounting (conditioning) for the effect from its own current state

$$\begin{aligned}
\mathrm{TE}_{X\to Y} &= I(y_{t+1};\mathbf{x}_t|\mathbf{y}_t) \\
&= H(\mathbf{x}_t,\mathbf{y}_t) - H(y_{t+1},\mathbf{x}_t,\mathbf{y}_t) + H(y_{t+1},\mathbf{y}_t) - H(\mathbf{y}_t) \\
&= \sum p(y_{t+1},\mathbf{x}_t,\mathbf{y}_t) \log \frac{p(y_{t+1}|\mathbf{x}_t,\mathbf{y}_t)}{p(y_{t+1}|\mathbf{y}_t)}
\end{aligned}$$

Joint entropies (and distributions) can have high dimension!

Entropy estimates from nearest neighbors [Kraskov et al, 2004]

Nonlinear causality measures (direct and indirect)

Transfer Entropy (TE) [Schreiber, 2000]

Measure the effect of X on Y at one time step ahead, accounting (conditioning) for the effect from its own current state

$$
\begin{aligned}
\mathrm{TE}_{X\to Y} &= I(y_{t+1};\mathbf{x}_t|\mathbf{y}_t) \\
&= H(\mathbf{x}_t,\mathbf{y}_t) - H(y_{t+1},\mathbf{x}_t,\mathbf{y}_t) + H(y_{t+1},\mathbf{y}_t) - H(\mathbf{y}_t) \\
&= \sum p(y_{t+1},\mathbf{x}_t,\mathbf{y}_t)\log\frac{p(y_{t+1}|\mathbf{x}_t,\mathbf{y}_t)}{p(y_{t+1}|\mathbf{y}_t)}
\end{aligned}
$$

Joint entropies (and distributions) can have high dimension!

Entropy estimates from nearest neighbors [Kraskov et al, 2004]

TE is equivalent to GCI when the stochastic process of (X,Y) is Gaussian [Barnett et al, PRE 2009]

Nonlinear causality measures (direct)

driving system: X, response system: Y,
conditioning on system Z, $Z = \{Z_1, Z_2, \dots, Z_{K-2}\}$
join all $K-2$ z-reconstructed vectors: $\mathbf{Z}_t = [\mathbf{z}_{1,t}, \dots, \mathbf{z}_{K-2,t}]$

Partial Transfer Entropy (PTE) [Vakorin et al, 2009; Papana et al, 2012]

Measure the effect of X on Y at T times ahead, accounting (conditioning) for the effect from its own current state and the current state of the other variables except X.

$$\mathrm{PTE}_{X\to Y|Z} = I(y_{t+1}; \mathbf{x}_t | \mathbf{y}_t, \mathbf{Z}_t)$$
$$= H(\mathbf{x}_t, \mathbf{y}_t|\mathbf{Z}_t) - H(y_{t+1}, \mathbf{x}_t, \mathbf{y}_t|\mathbf{Z}_t) + H(y_{t+1}, \mathbf{y}_t|\mathbf{Z}_t) - H(\mathbf{y}_t|\mathbf{Z}_t)$$

Joint entropies (and distributions) can have very high dimension!

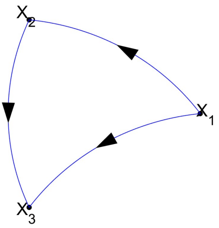
Example: Nonlinear stochastic process

Nonlinear stochastic map:

$$\begin{aligned}
x_{1,t} &= 3.4x_{1,t-1}(1-x_{1,t-1}^2)e^{-x_{1,t-1}^2} + 0.4e_{1,t} \\
x_{2,t} &= 3.4x_{2,t-1}(1-x_{2,t-1}^2)e^{-x_{2,t-1}^2} + 0.5{\color{red}x_{1,t-1}}x_{2,t-1} + 0.4e_{2,t} \\
x_{3,t} &= 3.4x_{3,t-1}(1-x_{3,t-1}^2)e^{-x_{3,t-1}^2} + 0.3{\color{red}x_{2,t-1}} + 0.5{\color{red}x_{1,t-1}^2} + 0.4e_{3,t}
\end{aligned}$$

[Model 7, Gourevich et al, 2006]

True connectivity network

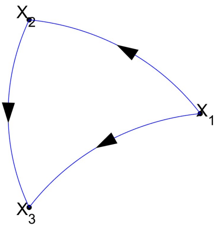


Nonlinear stochastic map:

$$
\begin{aligned}
x_{1,t} &= 3.4x_{1,t-1}(1-x^2_{1,t-1})e^{-x^2_{1,t-1}} + 0.4e_{1,t} \\
x_{2,t} &= 3.4x_{2,t-1}(1-x^2_{2,t-1})e^{-x^2_{2,t-1}} + 0.5x_{1,t-1}x_{2,t-1} + 0.4e_{2,t} \\
x_{3,t} &= 3.4x_{3,t-1}(1-x^2_{3,t-1})e^{-x^2_{3,t-1}} + 0.3x_{2,t-1} + 0.5x^2_{1,t-1} + 0.4e_{3,t}
\end{aligned}
$$

[Model 7, Gourevich et al, 2006]

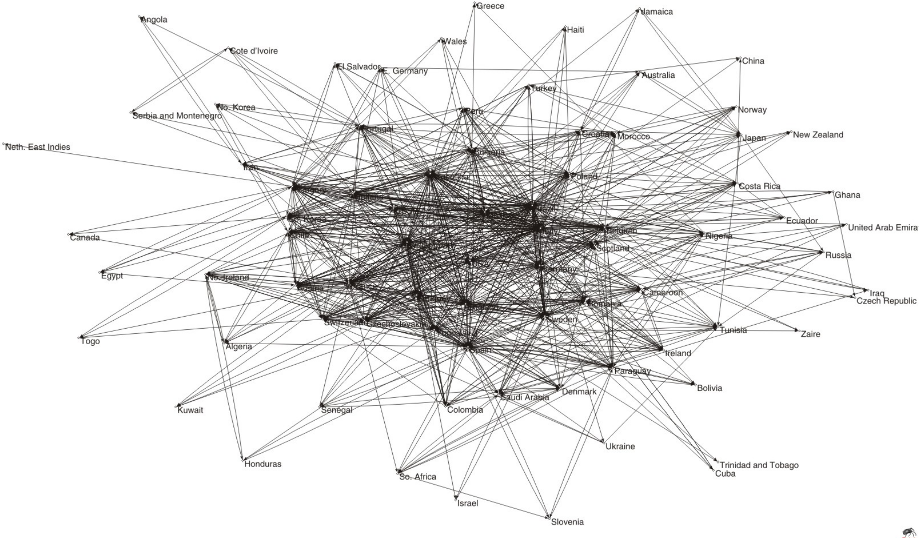
True connectivity network



Estimation of the correct causality effects from the time series?

1. Dependence measures in univariate time series
2. Interdependence in multivariate time series
3. Complex networks from multivariate time series
4. High-dimensional time series: Implications and solutions

Example: Games of world cup 1930 - 2006

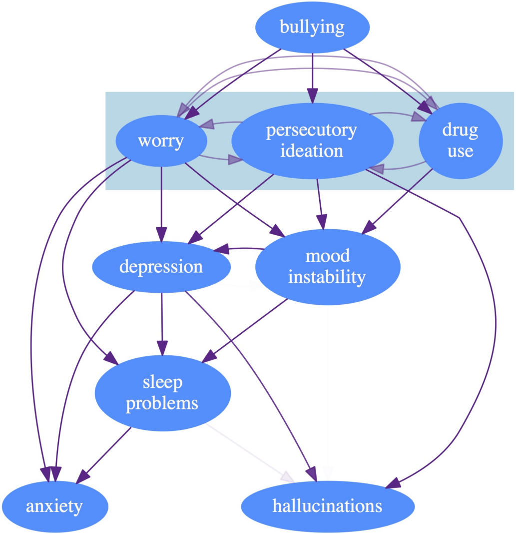


Example: Flight connections

Source: `https://au.pinterest.com/pin/488077678338752549/`

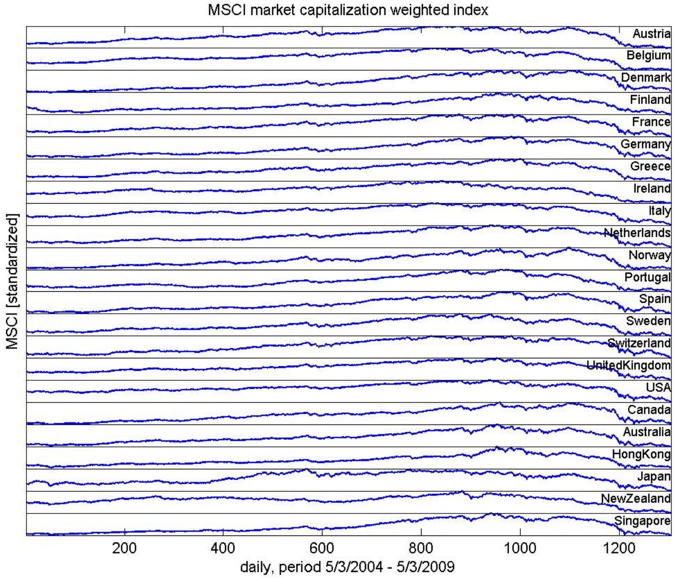
Example: Greek domestic flight connections

Example: Bullying causal effects



Source: `https://doi.org/10.1093/schbul/sbx013`

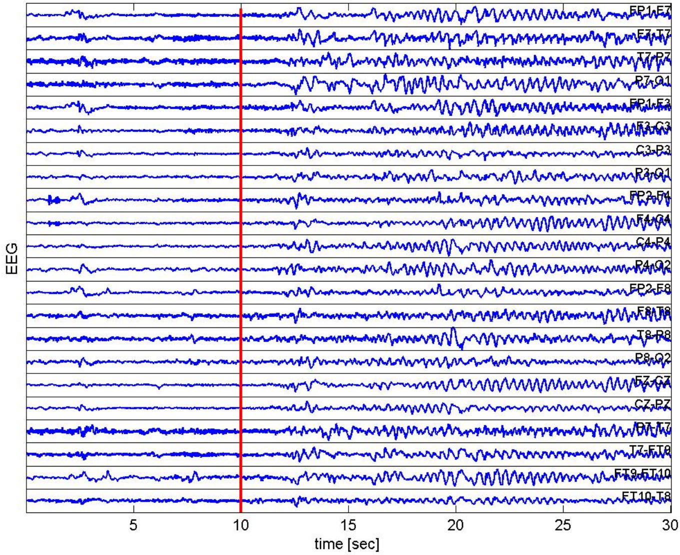
Example: Global financial market



Data source: `https://www.msci.com/market-cap-weighted-indexes`

Network ?

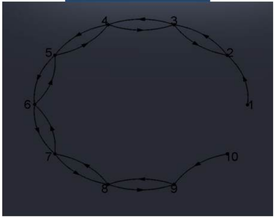
Example: Brain dynamical system



Data source: https://physionet.org/pn6/chbmit/chb08/

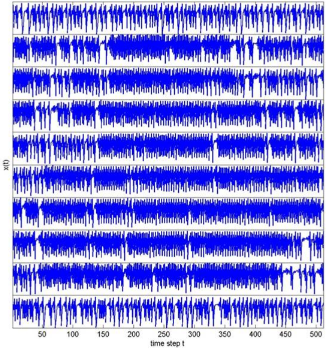
Network ?

[redacted]



Complex networks from multivariate time series

Multivariate time series



Causality measures

$X \rightarrow Y$, $X \rightarrow Y | Z$

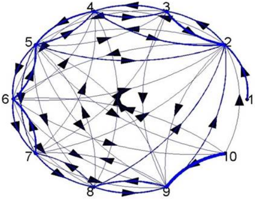
$r(X_t, Y_{t+\tau})$

$I(X_t, Y_{t+\tau})$

GCI, CGCI

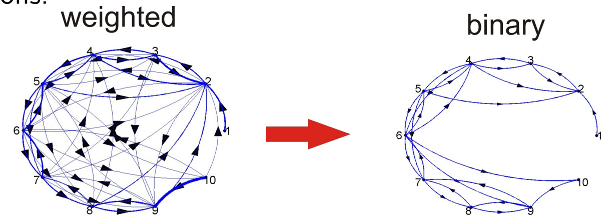
TE, PTE

Estimated network (weighted connections)



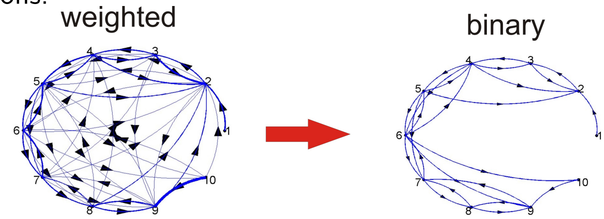
How to assess the presence of a connection?

Three possible ways to convert a network of weighted connections (the Granger causality measure) to a network of binary connections:



How to assess the presence of a connection?

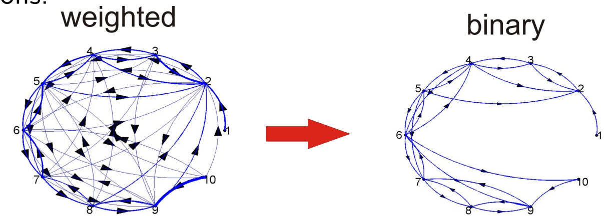
Three possible ways to convert a network of weighted connections (the Granger causality measure) to a network of binary connections:



1. Threshold on the measure magnitude, $q(i \to j) > \text{thr}$.

How to assess the presence of a connection?

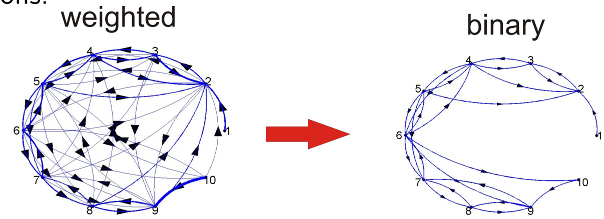
Three possible ways to convert a network of weighted connections (the Granger causality measure) to a network of binary connections:



1. Threshold on the measure magnitude, $q(i \to j) > \text{thr}$.
2. Threshold on the network density, only the $d\%$ largest $q(i \to j)$.

How to assess the presence of a connection?

Three possible ways to convert a network of weighted connections (the Granger causality measure) to a network of binary connections:



1. Threshold on the measure magnitude, $q(i \to j) > \text{thr}$.
2. Threshold on the network density, only the $d\%$ largest $q(i \to j)$.
3. Significance test on each $q(i \to j)$. Threshold, e.g. $\alpha = 0.05$ on the p-value of the test.
   Parametric or resampling test (resampling test for a nonlinear causality measure).

The problem of multiple testing

Significance resampling test on $q(i \rightarrow j)$ for each pair (X_i, X_j).

The problem of multiple testing

Significance resampling test on $q(i \to j)$ for each pair (X_i, X_j).

$K(K-1)$ significance tests $\Longrightarrow$ Correction for multiple testing

The problem of multiple testing

Significance resampling test on $q(i \to j)$ for each pair (X_i, X_j).

$K(K-1)$ significance tests $\Longrightarrow$ Correction for multiple testing

Popular choice:

False Discovery Rate (FDR) [Benjamini & Hochberg, 1995]

- $K(K-1)$ p-values in ascending order: $p_{(1)}, p_{(2)}, \ldots, p_{(K(K-1))}$
- Rejection for the k tests with $p \leq p_{(k)}$, where $p_{(k)}$ is the largest p-value for which $p_{(k)} < k\alpha/(K(K-1))$.

The problem of multiple testing

Significance resampling test on $q(i \to j)$ for each pair (X_i, X_j).

$K(K-1)$ significance tests $\Longrightarrow$ Correction for multiple testing

Popular choice:

False Discovery Rate (FDR) [Benjamini & Hochberg, 1995]

- $K(K-1)$ p-values in ascending order: $p_{(1)}, p_{(2)}, \dots, p_{(K(K-1))}$
- Rejection for the k tests with $p \leq p_{(k)}$, where $p_{(k)}$ is the largest p-value for which $p_{(k)} < k\alpha/(K(K-1))$.

Small p-value can only be obtained with large number of surrogates

The problem of multiple testing

Significance resampling test on $q(i \to j)$ for each pair (X_i, X_j).

$K(K-1)$ significance tests $\Longrightarrow$ Correction for multiple testing

Popular choice:

False Discovery Rate (FDR) [Benjamini & Hochberg, 1995]

- $K(K-1)$ p-values in ascending order: $p_{(1)}, p_{(2)}, \ldots, p_{(K(K-1))}$
- Rejection for the k tests with $p \leq p_{(k)}$, where $p_{(k)}$ is the largest p-value for which $p_{(k)} < k\alpha/(K(K-1))$.

Small p-value can only be obtained with large number of surrogates

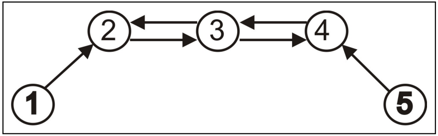
When K gets large, FDR requires **huge** M (impractical).

Example: coupled Henon maps

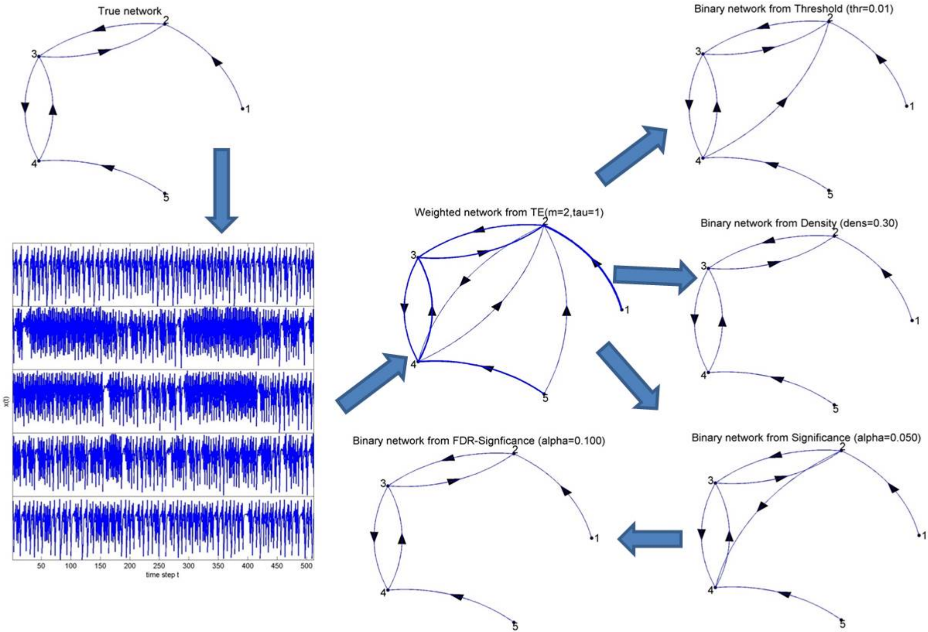
$$\begin{aligned} x_{1,t+1} &= 1.4 - x_{1,t}^2 + 0.3x_{1,t-1} \\ x_{i,t+1} &= 1.4 - (0.5C(x_{i-1,t} + x_{i+1,t}) + (1-C)x_{i,t})^2 + 0.3x_{i,t-1} \\ x_{K,t+1} &= 1.4 - x_{K,t}^2 + 0.3x_{K,t-1} \end{aligned}$$

C: coupling strength [Politi & Torcini, 1992]

Network structure for $K = 5$

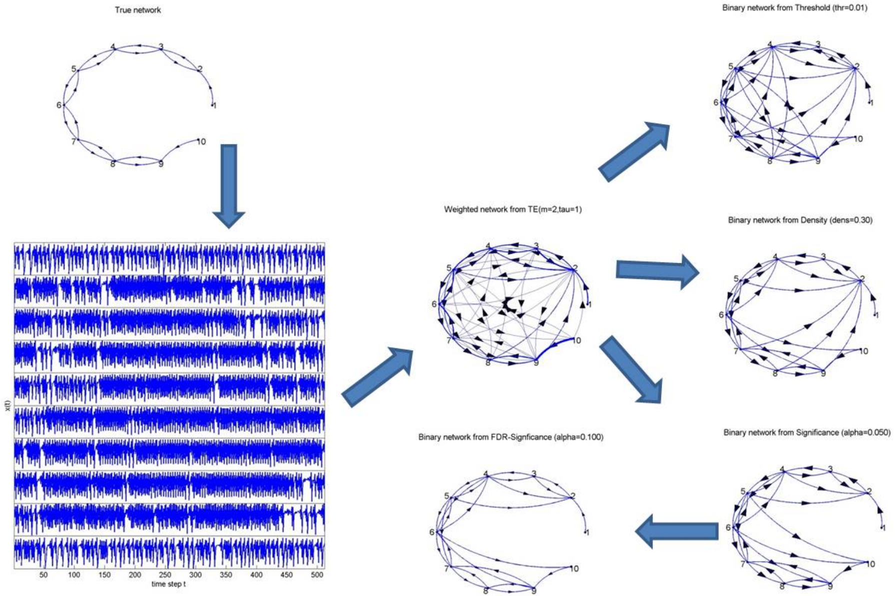


Example, TE, $K = 5$



1. Dependence measures in univariate time series
2. Interdependence in multivariate time series
3. Complex networks from multivariate time series
4. High-dimensional time series: Implications and solutions

Example, TE, $K = 10$



What if there are many observed variables?

The **curse of dimensionality**:

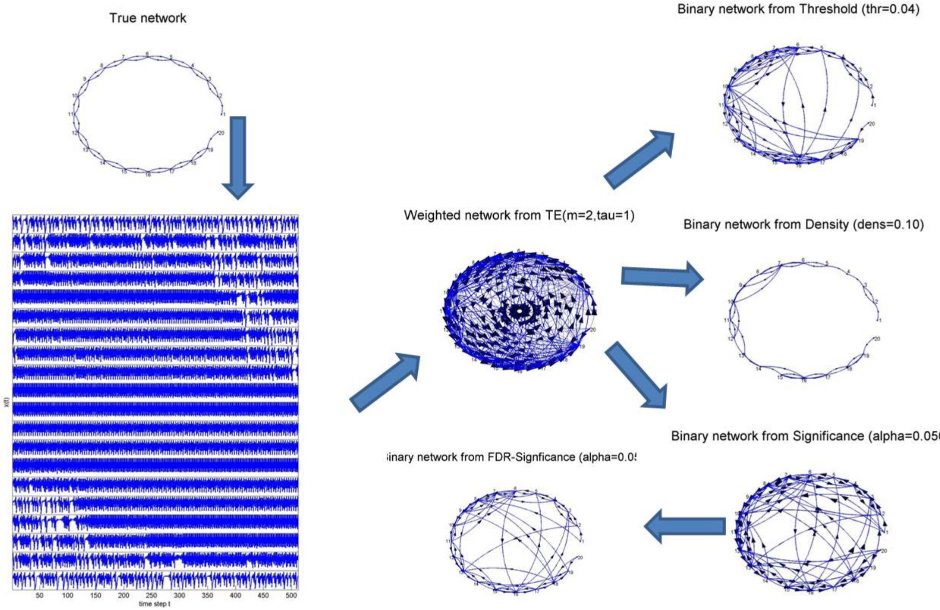
• For FDR, in general $M \sim K(K-1)/\alpha$. When K gets large, **huge** M may be required (impractical).

The **curse of dimensionality**:

• For FDR, in general $M \sim K(K-1)/\alpha$. When K gets large, **huge** M may be required (impractical).

• For $K > 2$, bivariate measures are likely to produce false couplings (indirect connections).

Example, TE, $K = 20$



What if there are many observed variables?

The **curse of dimensionality**:

• For FDR, in general $M \sim K(K-1)/\alpha$. When K gets large, **huge** M may be required (impractical).

• For $K > 2$, bivariate measures are likely to produce false couplings (indirect connections).

What if there are many observed variables?

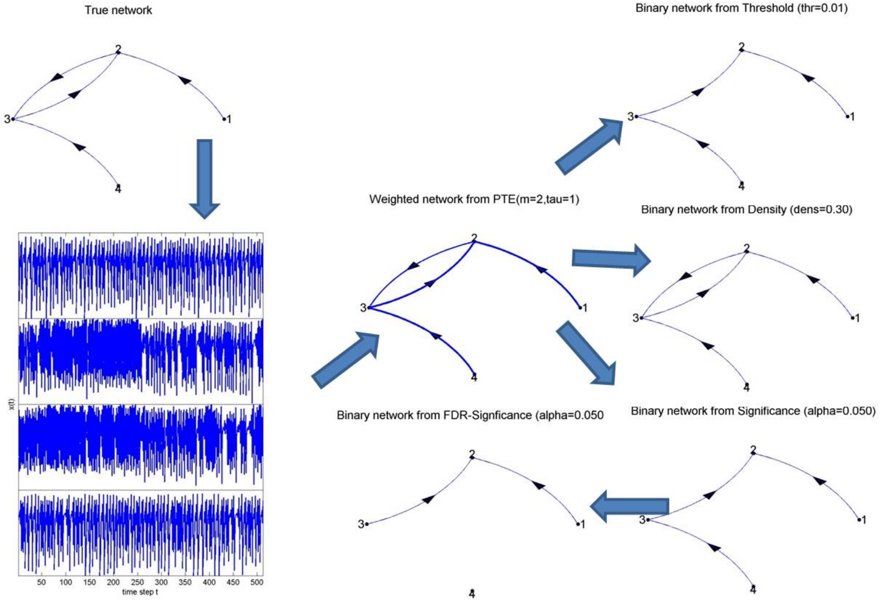
The **curse of dimensionality**:

• For FDR, in general $M \sim K(K-1)/\alpha$. When K gets large, **huge** M may be required (impractical).

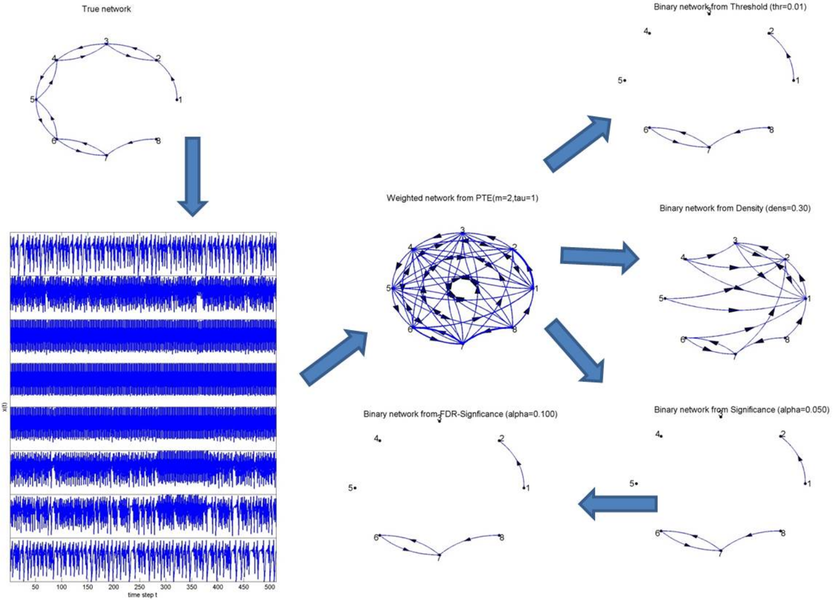
• For $K > 2$, bivariate measures are likely to produce false couplings (indirect connections).

• Multivariate measures require long time series, e.g. $\text{PTE}_{X \to Y|Z} = I(y_{t+1}; \mathbf{x}_t | \mathbf{y}_t, \mathbf{Z}_t)$ requires the estimation of entropy of $[y_{t+1}, \mathbf{x}_t, \mathbf{y}_t, \mathbf{Z}_t]'$ of dimension $1 + Km$.

Example, PTE, $K = 4$



Example, PTE, $K = 8$



Interdependence using Dimension Reduction

K time series $\{x_t, y_t\}_{t=1}^n$ and $\{\mathbf{z}_t\}_{t=1}^n = \{z_{1,t}, z_{2,t}, \ldots, z_{K-2,t}\}_{t=1}^n$
driving system: X, response system: Y,
conditioning on system Z, $Z = \{Z_1, Z_2, \ldots, Z_{K-2}\}$

If K large with respect to n multivariate Granger causality is problematic ("the curse of dimensionality").

Interdependence using Dimension Reduction

K time series $\{x_t, y_t\}_{t=1}^n$ and $\{\mathbf{z}_t\}_{t=1}^n = \{z_{1,t}, z_{2,t}, \ldots, z_{K-2,t}\}_{t=1}^n$
driving system: X, response system: Y,
conditioning on system Z, $Z = \{Z_1, Z_2, \ldots, Z_{K-2}\}$

If K large with respect to n multivariate Granger causality is problematic ("the curse of dimensionality").

In CGCI, the VAR model

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{p} b_i x_{t-i} + \sum_{i=1}^{p} A_i \mathbf{z}_{t-i} + e_{U,t}$$

has Kp lagged variables.

Interdependence using Dimension Reduction

K time series $\{x_t, y_t\}_{t=1}^n$ and $\{\mathbf{z}_t\}_{t=1}^n = \{z_{1,t}, z_{2,t}, \ldots, z_{K-2,t}\}_{t=1}^n$
driving system: X, response system: Y,
conditioning on system Z, $Z = \{Z_1, Z_2, \ldots, Z_{K-2}\}$

If K large with respect to n multivariate Granger causality is problematic ("the curse of dimensionality").

In CGCI, the VAR model

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{p} b_i x_{t-i} + \sum_{i=1}^{p} A_i \mathbf{z}_{t-i} + e_{U,t}$$

has Kp lagged variables.

Statistics methodology: dimension reduction, sparse regression, restricted regression, and sparse/retricted VAR models.

PMIME [Kugiumtzis, 2013]

Partial Mutual Information from Mixed Embedding (PMIME) applies dimension reduction using mutual information. The idea [Vlachos & Kugiumtzis, 2010]:

PMIME [Kugiumtzis, 2013]

Partial Mutual Information from Mixed Embedding (PMIME) applies dimension reduction using mutual information. The idea [Vlachos & Kugiumtzis, 2010]:

1. Find a subset $\mathbf{w}_y$ of lagged variables from X, Y and Z that explains best the future of Y.

PMIME [Kugiumtzis, 2013]

Partial Mutual Information from Mixed Embedding (PMIME) applies dimension reduction using mutual information. The idea [Vlachos & Kugiumtzis, 2010]:

1. Find a subset $\mathbf{w}_y$ of lagged variables from X, Y and Z that explains best the future of Y.
2. Quantify the information on Y ahead that is explained by the X-components in this subset.

Partial Mutual Information from Mixed Embedding (PMIME) applies dimension reduction using mutual information. The idea [Vlachos & Kugiumtzis, 2010]:

1. Find a subset $\mathbf{w}_y$ of lagged variables from X, Y and Z that explains best the future of Y.
2. Quantify the information on Y ahead that is explained by the X-components in this subset.

If there are no components of X in $\mathbf{w}_y$, then PMIME $= 0$.

PMIME [Kugiumtzis, 2013]

Partial Mutual Information from Mixed Embedding (PMIME) applies dimension reduction using mutual information. The idea [Vlachos & Kugiumtzis, 2010]:

1. Find a subset $\mathbf{w}_y$ of lagged variables from X, Y and Z that explains best the future of Y.
2. Quantify the information on Y ahead that is explained by the X-components in this subset.

If there are no components of X in $\mathbf{w}_y$, then PMIME $= 0$.

Similar approaches based on this idea: [Faes et al, PRE 2011; Stamaglia et al, PRE 2012; Runge et al, PRL 2012; Wibral et al, PLOSOne 2013; Runge et al, PRE 2015; edited book of Wibral, Vicente and Lizier "Directed information measures in Neuroscience", Springer, 2014.]

PMIME - 2

The mixed embedding scheme

- Start with an empty embedding vector $\mathbf{w}_t^0$, future vector of Y, y_{t+1}, and maximum lag L (or L_x for X, L_y for Y etc)
  $\mathbf{W}_t = \{x_t, \dots, x_{t-L-1}, y_t, \dots, y_{t-L-1}, z_{1,t}, \dots, z_{K-2,t-L-1}\}$

PMIME - 2

The mixed embedding scheme

- Start with an empty embedding vector $\mathbf{w}_t^0$, future vector of Y, y_{t+1}, and maximum lag L (or L_x for X, L_y for Y etc)
  $\mathbf{W}_t = \{x_t, \ldots, x_{t-L-1}, y_t, \ldots, y_{t-L-1}, z_{1,t}, \ldots, z_{K-2,t-L-1}\}$
- First embedding cycle: $w_t^1 = \text{argmax}_{w \in W_t} I(y_{t+1}; w)$, and $\mathbf{w}_t^1 = (w_t^1)$

PMIME - 2

The mixed embedding scheme

- Start with an empty embedding vector $\mathbf{w}_t^0$, future vector of Y, y_{t+1}, and maximum lag L (or L_x for X, L_y for Y etc)
  $\mathbf{W}_t = \{x_t, \ldots, x_{t-L-1}, y_t, \ldots, y_{t-L-1}, z_{1,t}, \ldots, z_{K-2,t-L-1}\}$
- First embedding cycle: $w_t^1 = \text{argmax}_{w \in W_t} I(y_{t+1}; w)$, and $\mathbf{w}_t^1 = (w_t^1)$
- At embedding cycle j suppose $\mathbf{w}_t^{j-1} = (w_t^1, w_t^2, \ldots, w_t^{j-1})$.

PMIME - 2

The mixed embedding scheme

- Start with an empty embedding vector $\mathbf{w}_t^0$, future vector of Y, y_{t+1}, and maximum lag L (or L_x for X, L_y for Y etc) $\mathbf{W}_t = \{x_t, \ldots, x_{t-L-1}, y_t, \ldots, y_{t-L-1}, z_{1,t}, \ldots, z_{K-2,t-L-1}\}$
- First embedding cycle: $w_t^1 = \text{argmax}_{w \in W_t} I(y_{t+1}; w)$, and $\mathbf{w}_t^1 = (w_t^1)$
- At embedding cycle j suppose $\mathbf{w}_t^{j-1} = (w_t^1, w_t^2, \ldots, w_t^{j-1})$. Add $w_t^j \in \mathbf{W}_t \setminus \mathbf{w}_t^{j-1}$ that maximizes mutual information to y_{t+1} conditioning on the current $\mathbf{w}_t^{j-1}$,

The mixed embedding scheme

- Start with an empty embedding vector $\mathbf{w}_t^0$, future vector of Y, y_{t+1}, and maximum lag L (or L_x for X, L_y for Y etc)
  $\mathbf{W}_t = \{x_t, \ldots, x_{t-L-1}, y_t, \ldots, y_{t-L-1}, z_{1,t}, \ldots, z_{K-2,t-L-1}\}$
- First embedding cycle: $w_t^1 = \text{argmax}_{w \in W_t} I(y_{t+1}; w)$, and $\mathbf{w}_t^1 = (w_t^1)$
- At embedding cycle j suppose $\mathbf{w}_t^{j-1} = (w_t^1, w_t^2, \ldots, w_t^{j-1})$.
  Add $w_t^j \in \mathbf{W}_t \setminus \mathbf{w}_t^{j-1}$ that maximizes mutual information to y_{t+1} conditioning on the current $\mathbf{w}_t^{j-1}$,
  $$w_t^j = \text{argmax}_{w \in mathbfW_t \setminus \mathbf{w}_t^{j-1}} I(y_{t+1}; w | \mathbf{w}_t^{j-1})$$

PMIME - 2

The mixed embedding scheme

- Start with an empty embedding vector $\mathbf{w}_t^0$, future vector of Y, y_{t+1}, and maximum lag L (or L_x for X, L_y for Y etc)
  $\mathbf{W}_t = \{x_t, \ldots, x_{t-L-1}, y_t, \ldots, y_{t-L-1}, z_{1,t}, \ldots, z_{K-2,t-L-1}\}$
- First embedding cycle: $w_t^1 = \text{argmax}_{w \in W_t} I(y_{t+1}; w)$, and $\mathbf{w}_t^1 = (w_t^1)$
- At embedding cycle j suppose $\mathbf{w}_t^{j-1} = (w_t^1, w_t^2, \ldots, w_t^{j-1})$.
  Add $w_t^j \in \mathbf{W}_t \setminus \mathbf{w}_t^{j-1}$ that maximizes mutual information to y_{t+1} conditioning on the current $\mathbf{w}_t^{j-1}$,
  $$w_t^j = \text{argmax}_{w \in mathbfW_t \setminus \mathbf{w}_t^{j-1}} I(y_{t+1}; w | \mathbf{w}_t^{j-1})$$
- Progressive vector building stops at step j ($\mathbf{w}_t = \mathbf{w}_t^{j-1}$):
  Criterion of hard threshold:
  $I\left(y_{t+1}; \mathbf{w}_t^{j-1}\right) / I\left(y_{t+1}; \mathbf{w}_t^{j}\right) > A$ (here $A = 0.95$)
  Criterion of adaptive threshold:
  randomization significance test on $I(y_{t+1}; w_t^j | \mathbf{w}_t^{j-1})$

PMIME - 3

The non-uniform mixed embedding vector of lags of all X, Y, Z for explaining y_{t+1}:

$$\mathbf{w}_t = (\underbrace{x_{t-\tau_{x1}}, \ldots, x_{t-\tau_{xm_x}}}_{\mathbf{w}_t^x}, \underbrace{y_{t-\tau_{y1}}, \ldots, y_{t-\tau_{ym_y}}}_{\mathbf{w}_t^y}, \underbrace{z_{t-\tau_{z1}}, \ldots, z_{t-\tau_{zm_z}}}_{\mathbf{w}_t^z})$$

PMIME - 3

The non-uniform mixed embedding vector of lags of all X, Y, Z for explaining y_{t+1}:

$$\mathbf{w}_t = (\underbrace{x_{t-\tau_{x1}}, \ldots, x_{t-\tau_{xm_x}}}_{\mathbf{w}_t^x}, \underbrace{y_{t-\tau_{y1}}, \ldots, y_{t-\tau_{ym_y}}}_{\mathbf{w}_t^y}, \underbrace{z_{t-\tau_{z1}}, \ldots, z_{t-\tau_{zm_z}}}_{\mathbf{w}_t^z})$$

The causality measure PMIME

$$R_{X\to Y|Z} = \frac{I(y_{t+1}; \mathbf{w}_t^x \mid \mathbf{w}_t^y, \mathbf{w}_t^Z)}{I(y_{t+1}; \mathbf{w}_t)}$$

- $R_{X\to Y|Z}$: information on the future of Y explained only by X-components of the embedding vector (given the components of Y and Z), normalized with the mutual information of the future of Y and the embedding vector.

PMIME - 3

The non-uniform mixed embedding vector of lags of all X, Y, Z for explaining y_{t+1}:

$$\mathbf{w}_t = (\underbrace{x_{t-\tau_{x1}}, \ldots, x_{t-\tau_{xm_x}}}_{\mathbf{w}_t^x}, \underbrace{y_{t-\tau_{y1}}, \ldots, y_{t-\tau_{ym_y}}}_{\mathbf{w}_t^y}, \underbrace{z_{t-\tau_{z1}}, \ldots, z_{t-\tau_{zm_z}}}_{\mathbf{w}_t^z})$$

The causality measure PMIME

$$R_{X\to Y|Z} = \frac{I(y_{t+1}; \mathbf{w}_t^x \mid \mathbf{w}_t^y, \mathbf{w}_t^Z)}{I(y_{t+1}; \mathbf{w}_t)}$$

- $R_{X\to Y|Z}$: information on the future of Y explained only by X-components of the embedding vector (given the components of Y and Z), normalized with the mutual information of the future of Y and the embedding vector.
- If $\mathbf{w}_t$ contains no components from X, then $R_{X\to Y|Z} = 0$ and X has no direct effect on the future of Y.

Three main advantages of PMIME

Three main advantages of PMIME

- $R_{X \to Y|Z} = 0$ when no significant causality is present, and $R_{X \to Y|Z} > 0$ when it is present
  [no significance test, no issues with multiple testing!]

PMIME - 4

Three main advantages of PMIME

- $R_{X \to Y|Z} = 0$ when no significant causality is present, and $R_{X \to Y|Z} > 0$ when it is present
  [no significance test, no issues with multiple testing!]
- mixed embedding for all variables is formed as part of the measure [it does not require the determination of embedding parameters for each variable]

PMIME - 4

Three main advantages of PMIME

- $R_{X\to Y|Z} = 0$ when no significant causality is present, and $R_{X\to Y|Z} > 0$ when it is present
  [no significance test, no issues with multiple testing!]
- mixed embedding for all variables is formed as part of the measure [it does not require the determination of embedding parameters for each variable]
- inclusion of more confounding variables only slows the computation and has no effect on statistical accuracy
  [**no "curse of dimensionality"** for any dimension of Z, only slow computation time]

Three main advantages of PMIME

- $R_{X\to Y|Z} = 0$ when no significant causality is present, and $R_{X\to Y|Z} > 0$ when it is present
  [no significance test, no issues with multiple testing!]
- mixed embedding for all variables is formed as part of the measure [it does not require the determination of embedding parameters for each variable]
- inclusion of more confounding variables only slows the computation and has no effect on statistical accuracy
  [**no "curse of dimensionality"** for any dimension of Z, only slow computation time]

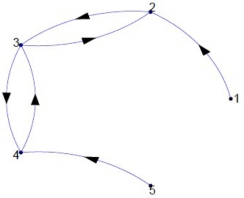
$\Rightarrow$ good candidate for causality analysis with many variables

Example: coupled Mackey-Glass

Coupled identical Mackey-Glass delayed differential equations

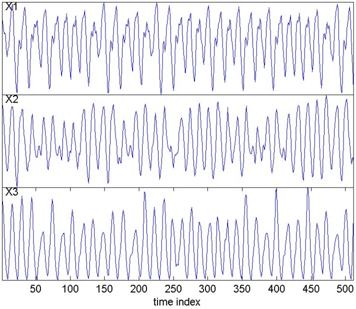
$$\dot{x}_i(t) = -0.1x_i(t) + \sum_{j=1}^{K} \frac{C_{ij}x_j(t-\Delta)}{1+x_j(t-\Delta)^{10}} \quad \text{for} \quad i = 1,\dots,K$$

$K = 5$



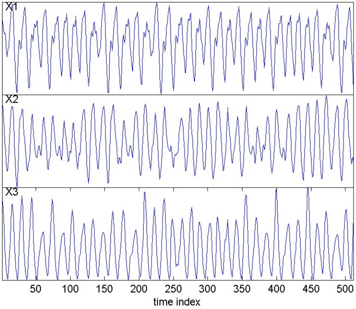
Mackey-Glass, $C = 0.2$

$\Delta = 20$

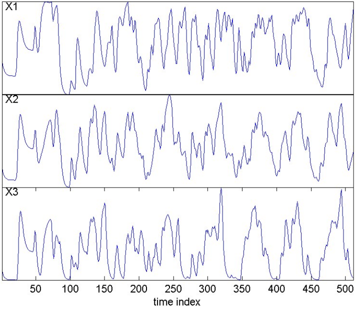


Mackey-Glass, $C = 0.2$

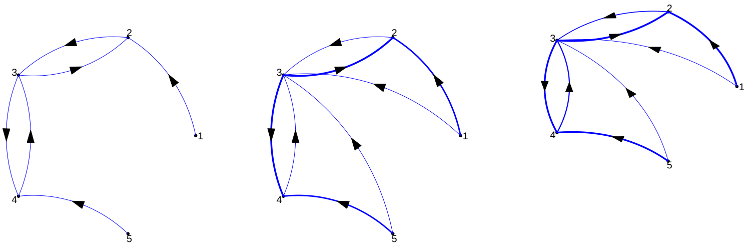
$\Delta = 20$



$\Delta = 100$

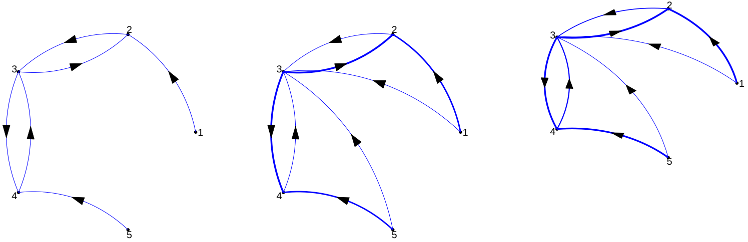


$K = 5$ True from PMIME ($\Delta = 20$) from PMIME ($\Delta = 100$)

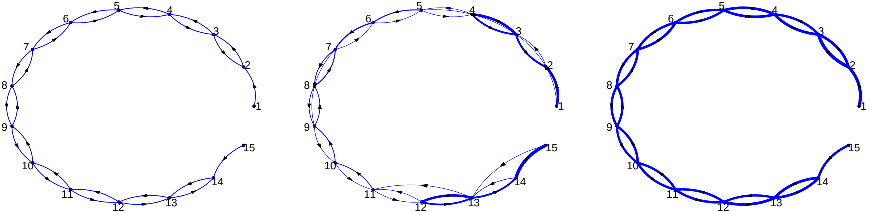


Mackey-Glass: true/estimated network [Kugiumtzis & Kimiskidis, 2015]

$K = 5$ True from PMIME ($\Delta = 20$) from PMIME ($\Delta = 100$)

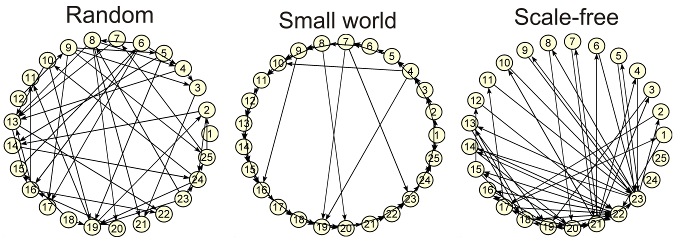


$K = 15$ True from PMIME ($\Delta = 20$) from PMIME ($\Delta = 100$)



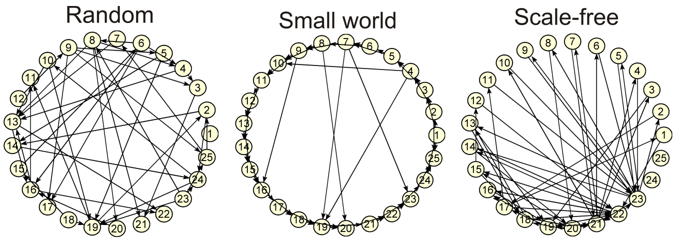
Can different network structures be detected?

Simulation: three types of networks for the generating system



Can different network structures be detected?

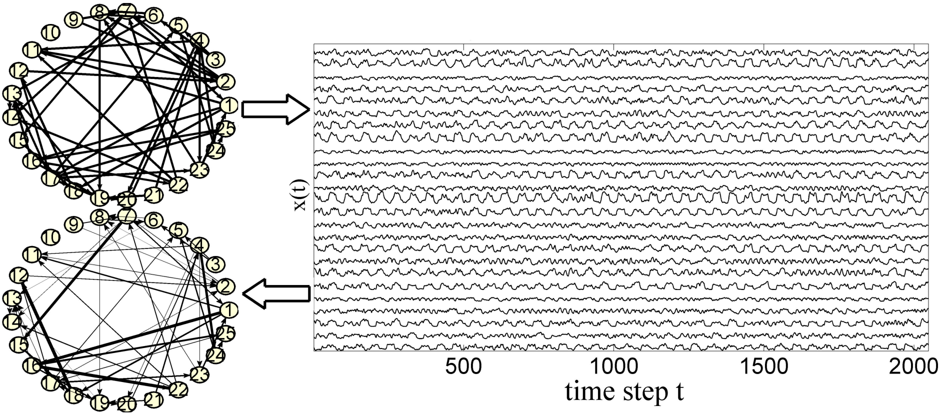
Simulation: three types of networks for the generating system



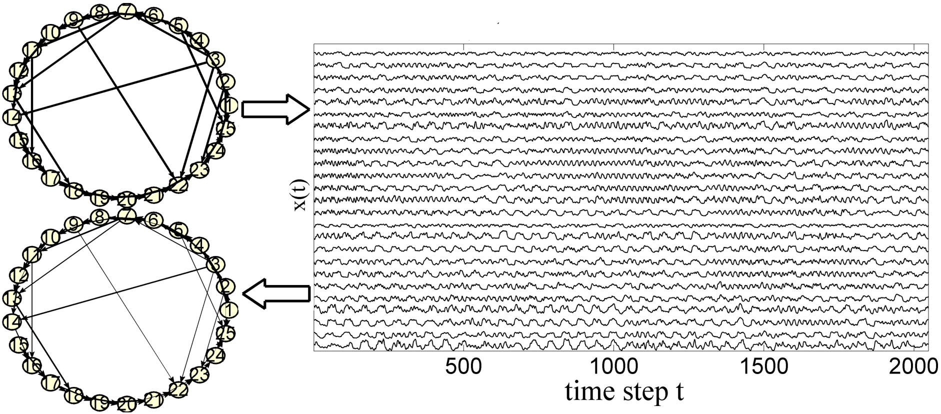
Generating system:
coupled Mackey-Glass system, $K = 25$, $\Delta = 100$, $C = 0.2$
with coupling structure defined by the network type

Causality measure: PMIME

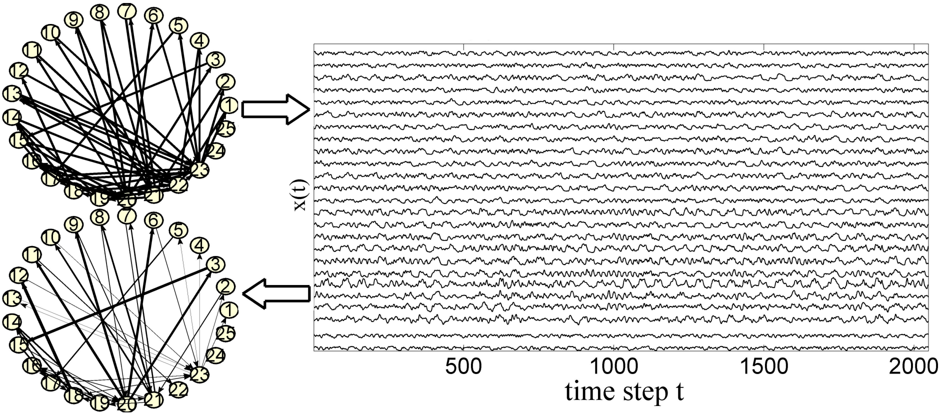
Estimation of the Random Network



Estimation of the Small-World Network



Estimation of the Scale-Free Network



Simulation example:

The network structure undergoes structural change at specific time points:
Random $\Rightarrow$ Small-World $\Rightarrow$ Scale-Free

Structural Change

Simulation example:

The network structure undergoes structural change at specific time points:
Random $\Rightarrow$ Small-World $\Rightarrow$ Scale-Free

Estimation of networks with PMIME at sliding windows

Simulation example:

The network structure undergoes structural change at specific time points:
Random $\Rightarrow$ Small-World $\Rightarrow$ Scale-Free

Estimation of networks with PMIME at sliding windows

Estimation of network characteristics on the PMIME networks

Simulation example:

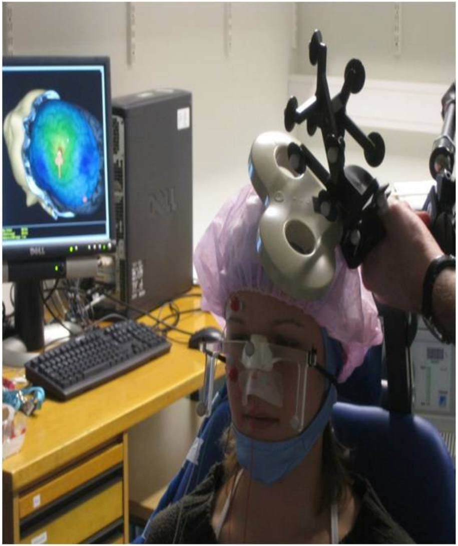
The network structure undergoes structural change at specific time points:
Random $\Rightarrow$ Small-World $\Rightarrow$ Scale-Free

Estimation of networks with PMIME at sliding windows

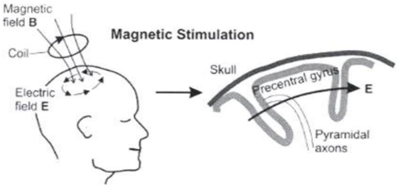
Estimation of network characteristics on the PMIME networks

Structural change detection,
[Slow], **[Middle]**, **[Fast]**, **[Very fast]**

EEG and Transcranial Magnetic Stimulation (TMS)



Jointly with Vassilis Kimiskidis, Medical School, AUTh



How does TMS act on epileptic brain connectivity?

How does TMS act on epileptic brain connectivity?

$\{X_1, X_2, \ldots, X_K\}$: K EEG channels, each represents a (sub)system

Practical problems to overcome:

TMS-EEG: brain connectivity analysis

How does TMS act on epileptic brain connectivity?

$\{X_1, X_2, \ldots, X_K\}$: K EEG channels, each represents a (sub)system

Practical problems to overcome:

- Application on small time windows $\Rightarrow$ limited data size

TMS-EEG: brain connectivity analysis

How does TMS act on epileptic brain connectivity?

$\{X_1, X_2, \ldots, X_K\}$: K EEG channels, each represents a (sub)system

Practical problems to overcome:

- Application on small time windows $\Rightarrow$ limited data size
- scalp EEG $\Rightarrow$ many channels $\Rightarrow$ many variables in Z to account for

How does TMS act on epileptic brain connectivity?

$\{X_1, X_2, \ldots, X_K\}$: K EEG channels, each represents a (sub)system

Practical problems to overcome:

- Application on small time windows $\Rightarrow$ limited data size
- scalp EEG $\Rightarrow$ many channels $\Rightarrow$ many variables in Z to account for
- Brain system is complex: the connectivity measure has to deal with $\begin{cases} \text{high dimensionality} \\ \text{nonlinearity?} \\ \text{sensitivity on free parameters?} \end{cases}$

How does TMS act on epileptic brain connectivity?

$\{X_1, X_2, \ldots, X_K\}$: K EEG channels, each represents a (sub)system

Practical problems to overcome:

- Application on small time windows $\Rightarrow$ limited data size
- scalp EEG $\Rightarrow$ many channels $\Rightarrow$ many variables in Z to account for
- Brain system is complex: the connectivity measure has to deal with $\begin{cases} \text{high dimensionality} \\ \text{nonlinearity?} \\ \text{sensitivity on free parameters?} \end{cases}$

PMIME addresses all these problems!

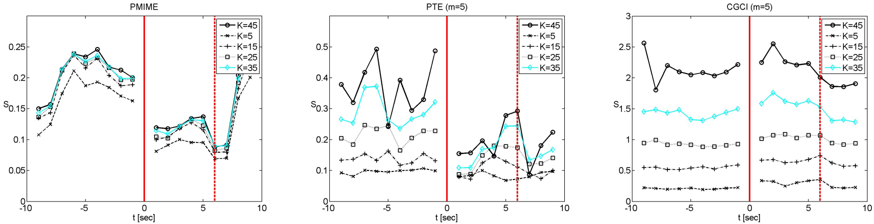
Example: compare PMIME to other measures on EEG

[Kugiumtzis, PRE, 2013]

One epileptiform discharge (ED) episode terminated by transcranial magnetic stimulation (TMS), totally 45 channels

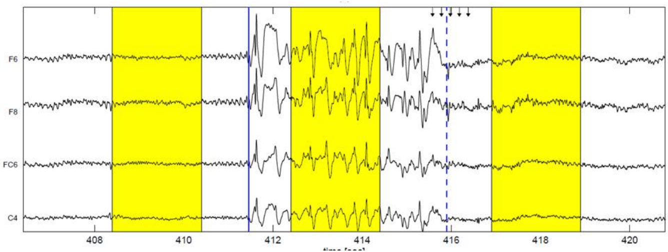
- Select randomly a subset of channels.
- Compute the connectivity measures on the subset at each sliding window
- Compute average connectivity strength at each sliding window.
- Repeat the steps above a number of times (here 12).

... for subsets of 5, 15, 25, 35 and once for 45 channels.



Preprocessing: **[One Episode]**

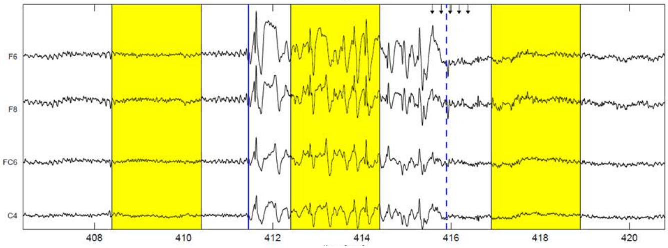
- replacement of TMS artifact, high order FIR
- rejection of channels with artifacts
- reference to infinity (REST) [Qin et al, ClinNeuroph 2010]
- overlapping windows of 2s, a sliding step of 0.5s

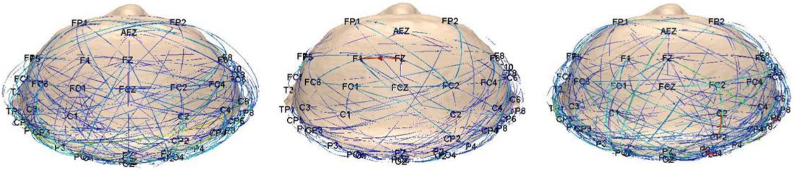


Epileptiform discharges induced by TMS

Preprocessing: **[One Episode]**

- replacement of TMS artifact, high order FIR
- rejection of channels with artifacts
- reference to infinity (REST) [Qin et al, ClinNeuroph 2010]
- overlapping windows of 2s, a sliding step of 0.5s



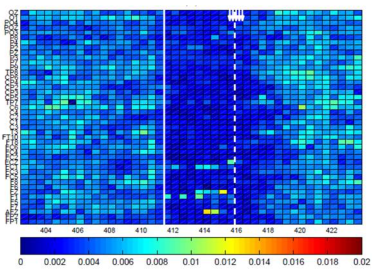


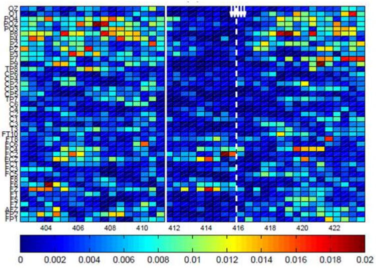
subject 1 with focal seizure, ED episode ends with TMS

In-Strength

Out-Strength

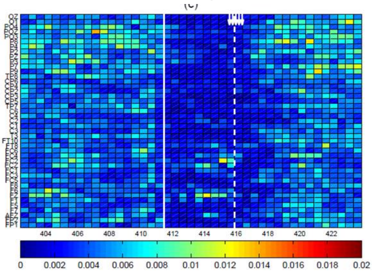
[Kugiumtzis & Kimiskidis, 2015]

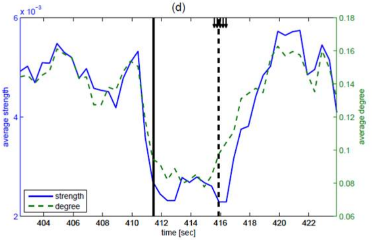




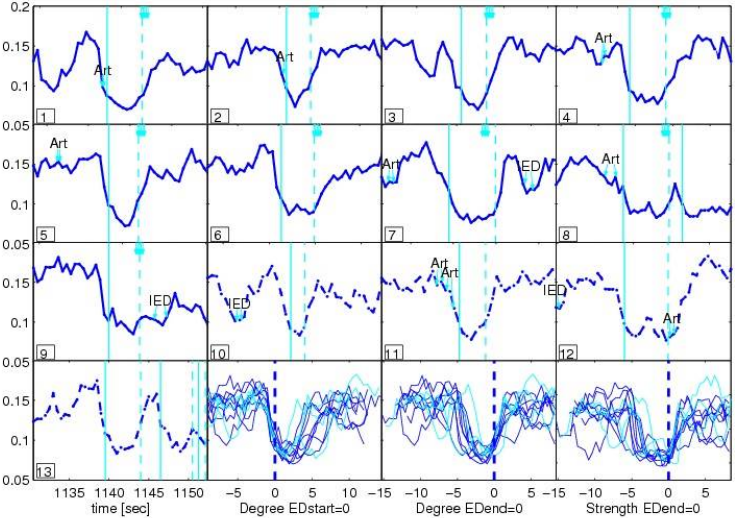
In-Out-Strength

Average strength / degree

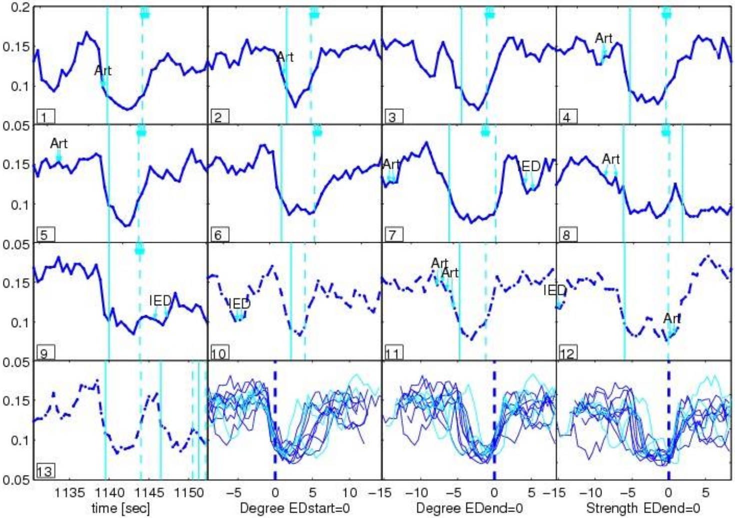




Subject 1 with focal seizure, 13 episodes, average degree

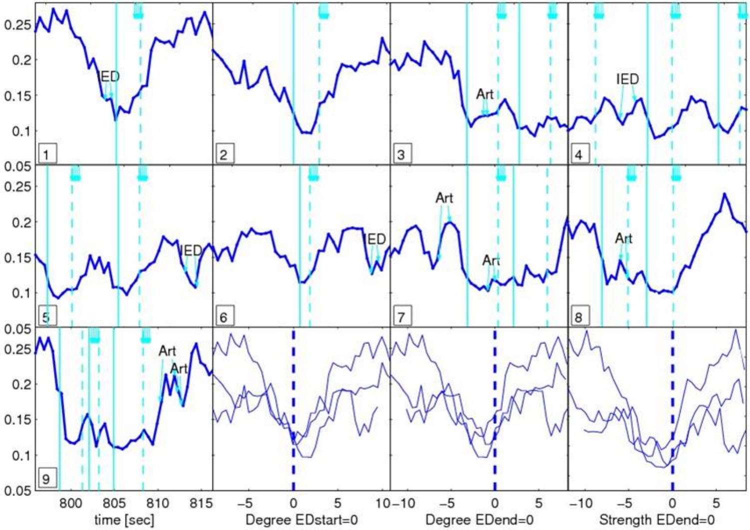


Subject 1 with focal seizure, 13 episodes, average degree



TMS terminates the ED prematurely and restores the network structure as if it would have terminated spontaneously

Subject 2 with focal seizure, 9 episodes, average degree



TMS terminates the ED prematurely and restores the network structure as if it would have terminated spontaneously

Many network indices (totally 78) computed on the PMIME-causality networks

Symbol	Description
deg^m	degree distribution, m=mean,std,skewness,kurtosis
str^m	strength distribution, m=mean,std,skewness,kurtosis
TrR_k	transitivity ratio, k=binary undirected (bu),binary directed (bd) weighted directed (wd)
$EigC^m$	eigenvector centrality distribution, m=mean,std
λ_k	characteristic path length, k=bd,wd
GE_k	global efficiency, k=bd,wd
ϵ_k^m	eccentricity distribution, m=mean,std and k=bd,wd
rad_k	radius, k=bd,wd
d_k	diameter, k=bd,wd
C_k^m	clustering coefficient distribution,m=mean,std and k=bd,wd
g_k^m	betweenness centrality distribution,m=mean,std and k=bd,wd
$e-g_k^m$	edge betweenness centrality distribution,m=mean,std and k=bd,wd
LE_k^m	local efficiency distribution,m=mean,std and k=bd,wd
$3motif(i)$	i^{th} motif of 3 nodes, i=1,2,...13
$modul(i)$	modularity for i modules, i=2,3,5
$r_{deg}(i,j)$	assortativity coefficient in terms of the degree, i=in,out and j=in,out or i,j=und
$r_{str}(i,j)$	assortativity coefficient in terms of the strength, i=in,out and j=in,out or i,j=und
p_{top}	Rent exponent:topological
p_{ph}	Rent exponent:physical
p_{ee}	Rent exponent:efficient embedding
SW_k	small-worldness, k=bd,wd
kcs	k-core size, k=90-percentile of degree distribution
scs	s-core size, k=90-percentile of strength distribution
ϕ_k	Rich club coefficient, k=bd,wd
$cycprob_1$	fraction of 3-cycles out of 3-paths
$cycprob_2$	probability: non-cyclic 2-path extend to 3-cycle

Many network indices (totally 78) computed on the PMIME-causality networks

Symbol	Description
deg^m	degree distribution, m=mean,std,skewness,kurtosis
str^m	strength distribution, m=mean,std,skewness,kurtosis
TrR_k	transitivity ratio, k=binary undirected (bu),binary directed (bd) weighted directed (wd)
$EigC^m$	eigenvector centrality distribution, m=mean,std
λ_k	characteristic path length, k=bd,wd
GE_k	global efficiency, k=bd,wd
ϵ_k^m	eccentricity distribution, m=mean,std and k=bd,wd
rad_k	radius, k=bd,wd
d_k	diameter, k=bd,wd
C_k^m	clustering coefficient distribution,m=mean,std and k=bd,wd
g_k^m	betweenness centrality distribution,m=mean,std and k=bd,wd
$e-g_k^m$	edge betweenness centrality distribution,m=mean,std and k=bd,wd
LE_k^m	local efficiency distribution,m=mean,std and k=bd,wd
$3motif(i)$	i^{th} motif of 3 nodes, i=1,2,...13
$modul(i)$	modularity for i modules, i=2,3,5
$r_{deg}(i,j)$	assortativity coefficient in terms of the degree, i=in,out and j=in,out or i,j=und
$r_{str}(i,j)$	assortativity coefficient in terms of the strength, i=in,out and j=in,out or i,j=und
p_{top}	Rent exponent:topological
p_{ph}	Rent exponent:physical
p_{ee}	Rent exponent:efficient embedding
SW_k	small-worldness, k=bd,wd
kcs	k-core size, k=90-percentile of degree distribution
scs	s-core size, k=90-percentile of strength distribution
ϕ_k	Rich club coefficient, k=bd,wd
$cycprob_1$	fraction of 3-cycles out of 3-paths
$cycprob_2$	probability: non-cyclic 2-path extend to 3-cycle

For both subjects and pairs
preED - ED
ED - postED

Measure	AUROC
deg^{mean}	0.9296
d_{bd}	0.9268
$3motif(1)$	0.9231
λ_{bd}	0.9231
str^{mean}	0.9207
$3motif(3)$	0.9206
$3motif(5)$	0.9199
LE_{bd}^{mean}	0.9173
GE_{wd}	0.9163
TrR_{wd}	0.9150

Many network indices (totally 78) computed on the PMIME-causality networks

Symbol	Description
deg^m	degree distribution, m=mean,std,skewness,kurtosis
str^m	strength distribution, m=mean,std,skewness,kurtosis
TrR_k	transitivity ratio, k=binary undirected (bu),binary directed (bd) weighted directed (wd)
$EigC^m$	eigenvector centrality distribution, m=mean,std
λ_k	characteristic path length, k=bd,wd
GE_k	global efficiency, k=bd,wd
ϵ_k^m	eccentricity distribution, m=mean,std and k=bd,wd
rad_k	radius, k=bd,wd
d_k	diameter, k=bd,wd
C_k^m	clustering coefficient distribution,m=mean,std and k=bd,wd
g_k^m	betweenness centrality distribution,m=mean,std and k=bd,wd
$e-g_k^m$	edge betweenness centrality distribution,m=mean,std and k=bd,wd
LE_k^m	local efficiency distribution,m=mean,std and k=bd,wd
$3motif(i)$	i[th] motif of 3 nodes, i=1,2,...13
$modul(i)$	modularity for i modules, i=2,3,5
$r_{deg}(i,j)$	assortativity coefficient in terms of the degree, i=in,out and j=in,out or i,j=und
$r_{str}(i,j)$	assortativity coefficient in terms of the strength, i=in,out and j=in,out or i,j=und
p_{top}	Rent exponent:topological
p_{ph}	Rent exponent:physical
p_{ee}	Rent exponent:efficient embedding
SW_k	small-worldness, k=bd,wd
kcs	k-core size, k=90-percentile of degree distribution
scs	s-core size, k=90-percentile of strength distribution
ϕ_k	Rich club coefficient, k=bd,wd
$cycprob_1$	fraction of 3-cycles out of 3-paths
$cycprob_2$	probability: non-cyclic 2-path extend to 3-cycle

For both subjects and pairs
preED - ED
ED - postED

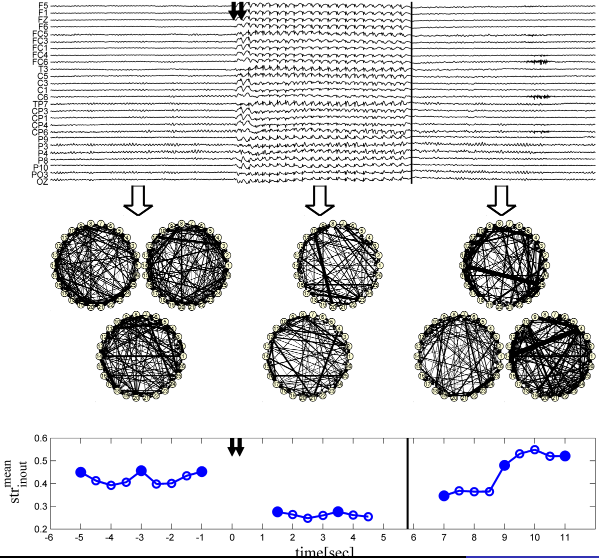
Measure	AUROC
deg^{mean}	0.9296
d_{bd}	0.9268
$3motif(1)$	0.9231
λ_{bd}	0.9231
str^{mean}	0.9207
$3motif(3)$	0.9206
$3motif(5)$	0.9199
LE_{bd}^{mean}	0.9173
GE_{wd}	0.9163
TrR_{wd}	0.9150

Many networks indices discriminate well preED, ED, postED

ED induced by TMS, **[PMIME on 2s windows]** [Kugiumtzis et al, 2016]

ED induced by TMS, **[PMIME on 2s windows]** [Kugiumtzis et al, 2016]

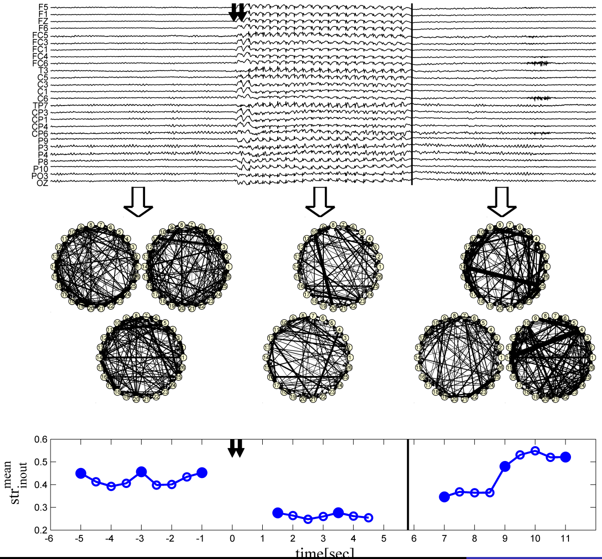
One episode



subject 1 with genetic generalized epilepsy (GGE)

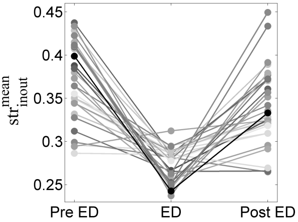
ED induced by TMS, **[PMIME on 2s windows]** [Kugiumtzis et al, 2016]

One episode

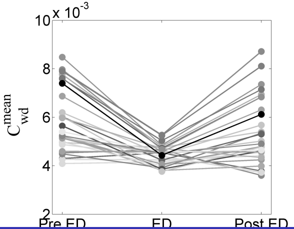


29 episodes

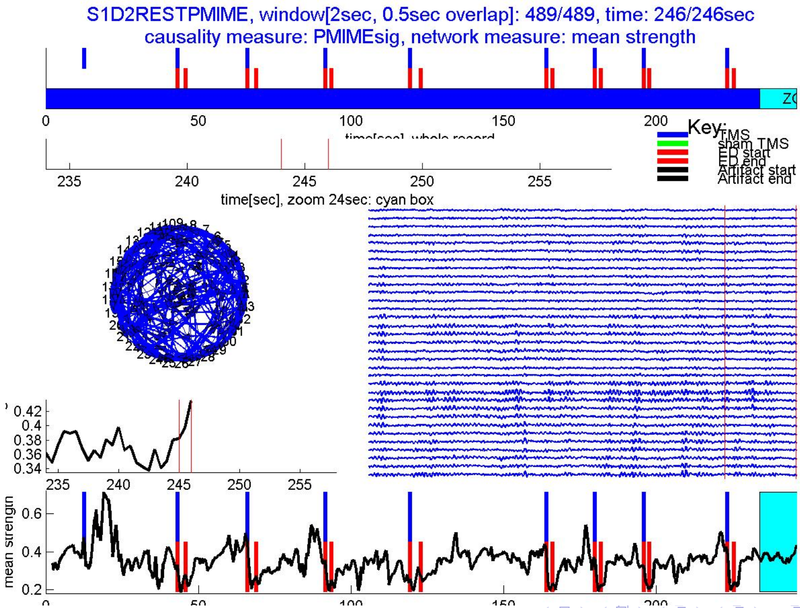
Average strength



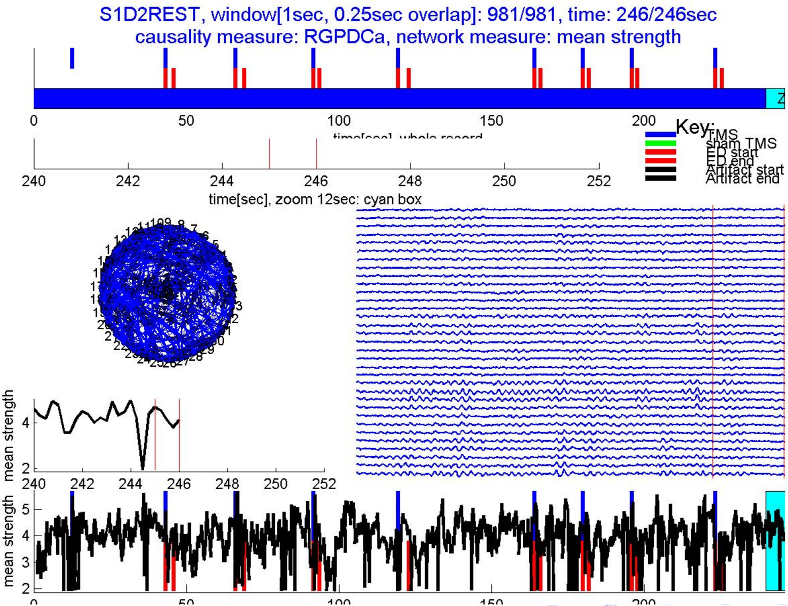
Average clustering coefficient



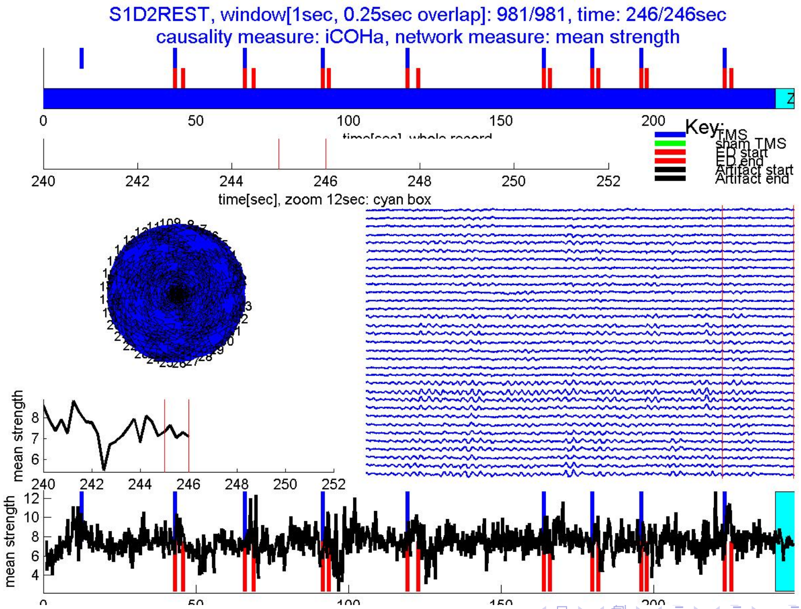
PMIME on 2s window



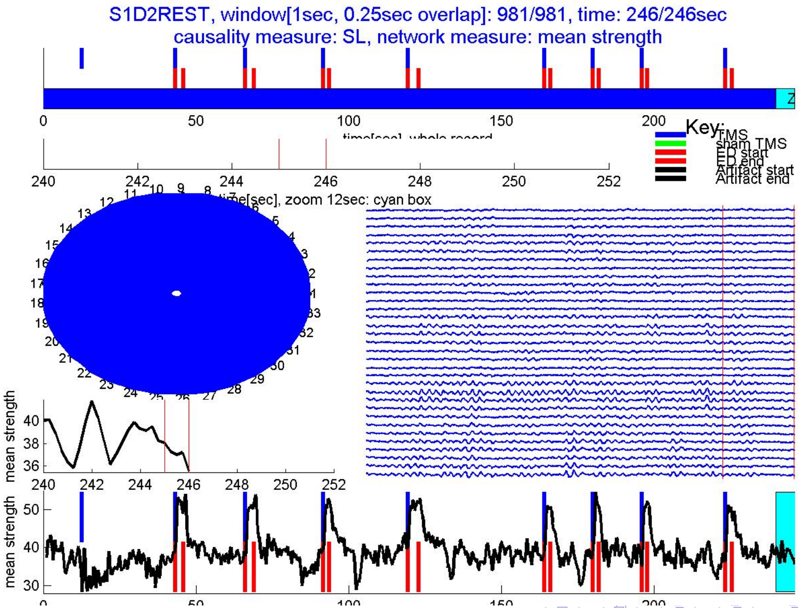
RGPDC(α) on 1s window



iCoh(α) on 1s window



Synchronization Likelihood (SL) on 1s window

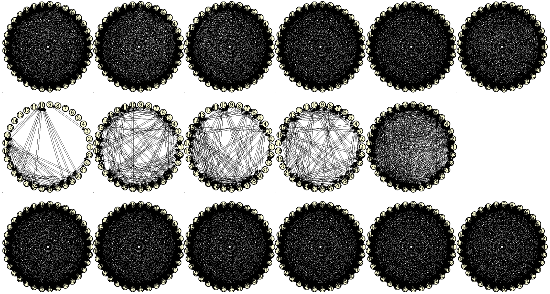


subject 1 with GGE

ED induced by TMS, SL on 1s windows,

ED induced by TMS, SL on 1s windows,

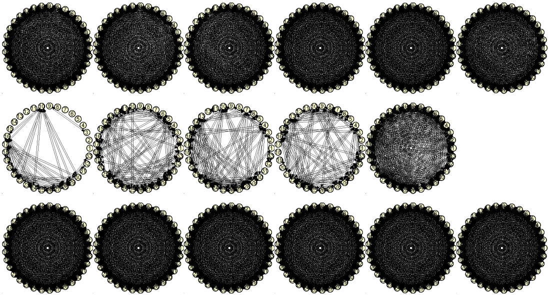
One episode



subject 1 with GGE

ED induced by TMS, SL on 1s windows,

One episode



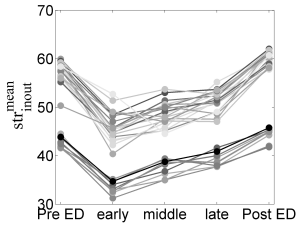
29 episodes

Average strength

SL

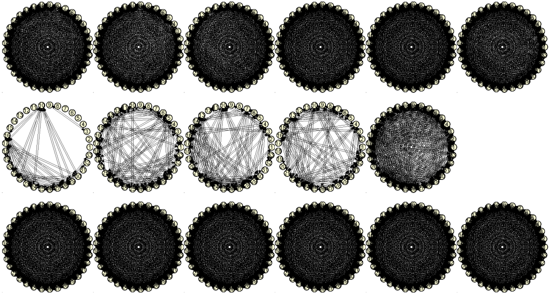
iCOH(α)

SL (other parameters)



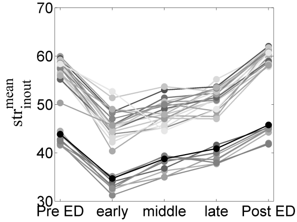
ED induced by TMS, SL on 1s windows,

One episode

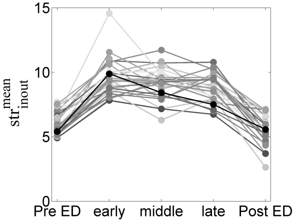


29 episodes Average strength

SL



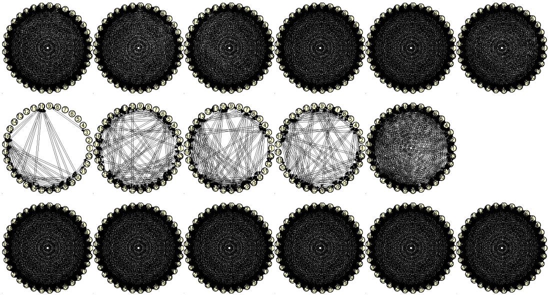
iCOH(α)



SL (other parameters)

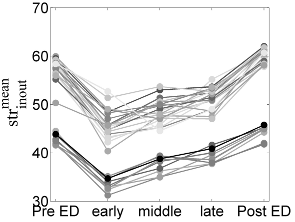
ED induced by TMS, SL on 1s windows,

One episode

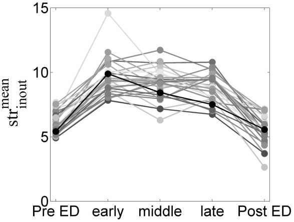


29 episodes Average strength

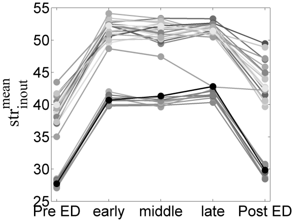
SL



iCOH(α)



SL (other parameters)



Summary

Summary

- "Focus on the reliable estimation of interdependence before you focus on complex networks"

Summary

- "Focus on the reliable estimation of interdependence before you focus on complex networks"
- "Bivariate measures are short for forming complex networks"

Summary

- "Focus on the reliable estimation of interdependence before you focus on complex networks"
- "Bivariate measures are short for forming complex networks"
- "Multivariate measures are hard to estimate"

Summary

- "Focus on the reliable estimation of interdependence before you focus on complex networks"
- "Bivariate measures are short for forming complex networks"
- "Multivariate measures are hard to estimate"
- PMIME is model-free and almost parameter-free, can estimate nonlinear direct causal effects in the presence of many variables

Summary

- "Focus on the reliable estimation of interdependence before you focus on complex networks"

- "Bivariate measures are short for forming complex networks"

- "Multivariate measures are hard to estimate"

- PMIME is model-free and almost parameter-free, can estimate nonlinear direct causal effects in the presence of many variables

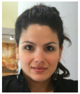
- How can we learn the underlying dynamics of high-dimensional time series?

Aristotle University of Thessaloniki

Collaborators:

collaborators in the research team

Alkis Tsimpiris
TEI Central Makedonia



Elsa Siggiridou
PhD candidate

Maria Papapetrou
PhD candidate

Akylas Fotiadis
PhD candidate

other collaborators

Vassilios K. Kimiskidis
Neurologist, Assoc.Prof.
Aristotle Uni

Paal G. Larsson
Neurophysiologist
Oslo Rikshospitalet

Aggeliki Papana
PostDoc
Macedonia Uni